

SEQUENCE LISTING

## (1) GENERAL INFORMATION:

(i) APPLICANT: Kalchman, Michael

Hayden, Michael R.

Hackam, Abigail

Chopra, Vikramjit Singh

Nicholson, Donald W.

Vallaincourt, John P.

Rasper, Dita M.

(ii) TITLE OF INVENTION: Apoptosis Modulators That Interact with the  
Huntington's Disease Gene

(iii) NUMBER OF SEQUENCES: 44

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Oppedahl &amp; Larson

(B) STREET: PO Box 5270

(C) CITY: Frisco

(D) STATE: CO

(E) COUNTRY: USA

(F) ZIP: 80443-5270

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Kb storage

(B) COMPUTER: IBM Compatible

(C) OPERATING SYSTEM: MS DOS 5.0

(D) SOFTWARE: WordPerfect

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:

(B) FILING DATE:

(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Larson, Marina T.

(B) REGISTRATION NUMBER: 32038

(C) REFERENCE/DOCKET NUMBER: UBC.P.013US2

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (970) 668-2050

(B) TELEFAX: (970) 668-2052

## (2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1164

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(ix) FEATURE: cDNA for Huntingtin-interacting protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```

ACAGCTGACA CCCTGCAAGG CCACCGGGAC CGCTTCATGG AGCAGTTTAC 50
AAAGTTGAAA GATCTGTTCT ACCGCTCCAG CAACCTGCAG TACTTCAAGC 100
GGGTCATTCA GATCCCCCAG CTGCCTGAGA ACCCACCCTAA CTTCCTGCGA 150
GCCTCAGCCC TGTCAGAACA TATCAGCCCT GTGGTGGTGA TCCCTGCAGA 200
GGCCTCATCC CCCGACAGCG AGCCAGTCCT AGAGAAGGAT GACCTCATGG 250
ACATGGATGC CTCTCAGCAG AATTTATTTG ACAACAAGTT TGATGACNTC 300
TTTGGCAGTT CATCCAGCAG TGATCCCTTC AATTTCAACA GTCAAAATGG 350
TGTGAACAAG GATGAGAAGG ACCACTTAAT TGAGCGACTA TACAGAGAGA 400
TCAGTGGATT GAAGGCACAG CTAGAAAACA TGAAGACTGA GAGCCAGCGG 450
GTTGTGCTGC AGCTGAAGGG CCACGTCAGC GAGCTGGAAG CAGATCTGGC 500
CGAGCAGCAG CACCTGCGGC AGCAGGCGGC CGACGACTGT GAATTCCTGC 550
GGGCAGAACT GGACGAGCTC AGGNGGCAGC GGGAGGACAC CGAGAAGGCT 600
CAGCGGAGCC TGTCTGAGAT AGAAAGGAAA GCTCAAGCCA ATGAACAGCG 650
ATATAGCAAG CTAAAGGAGA AGTACAGCGA GCTGGTTCAG AACCACGCTG 700
ACCTGCTGCG GAAGAATGCA GAGGTGACCA AACAGGTGTC CATGGCCAGA 750
CAAGCCCAGG TAGATTTGGA ACGAGAGAAA AAAGAGCTGG AGGATTCGTT 800
GGAGCGCATC AGTGACCAGG GCCAGCGGAA GACTCAAGAA CAGCTGGAAG 850
TTCTAGAGAG CTTGAAGCAG GAACTTGGA CAAGCCAACG GGAGCTTCAG 900
GTTCTGCAAG GCAGCCTGGA AACTTCTGCC CAGTCAGAAG CAAACTGGGC 950
AGCCGAGTTC GCCGAGCTAG AGAAGGAGCG GGACAGCCTG GTGAGTGGCG 1000
CAGCTCATAG GGAGGAGGAA TTATCTGCTC TTCGAAAGA ACTGCAGGAC 1050
ACTCAGCTCA AACTGGCCAG CACAGAGGAA TCTATGTGCC AGCTTGCCAA 1100
AGACCAACGA AAAATGCTTC TGGTGGGGTC CAGGAAGGCT GCGGAGCAGG 1150
TGATACAAGA CGCG 1164

```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 386

(B) TYPE: protein

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: no

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(ix) FEATURE: Huntingtin-interacting protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```

Thr Ala Asp Thr Leu Gln Gly His Arg Asp Arg Phe Met Glu Gln
1           5           10           15

```

```

Phe Thr Lys Leu Lys Asp Leu Phe Tyr Arg Ser Ser Asn Leu Gln
                20           25           30

```

```

Tyr Phe Lys Arg Val Ile Gln Ile Pro Gln Leu Pro Glu Asn Pro

```

3

	290		295		300
Gln Gly Ser Leu Glu Thr Ser Ala Gln Ser Glu Ala Asn Trp Ala					
	305		310		315
Ala Glu Phe Ala Glu Leu Glu Lys Glu Arg Asp Ser Leu Val Ser					
	320		325		330
Gly Ala Ala His Arg Glu Glu Glu Leu Ser Ala Leu Arg Lys Glu					
	335		340		345
Leu Gln Asp Thr Gln Leu Lys Leu Ala Ser Thr Glu Glu Ser Met					
	350		355		360
Cys Gln Leu Ala Lys Asp Gln Arg Lys Met Leu Leu Val Gly Ser					
	365		370		375
Arg Lys Ala Ala Glu Gln Val Ile Gln Asp Ala					
	380		385 386		

## (2) INFORMATION FOR SEQ ID NO:3:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4796

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(ix) FEATURE: cDNA for Huntingtin-interacting protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```

CAGTGTACGG TTGATCATAT AACGCCGCGG GCGGGGATTG GTTTATATATAT 50
CGCAAATTGA TNTAGGGGGG GGGGGATGGN CAGAGATTTC GCTTCATTAG 100
GCCATTATAA GCAGGAAGGG TTTCAAGGAA AAAAACCCAG AAAGTGCATA 150
TTGCACCCAC CATGAGAAAG GGGCAACAGA CCTTNTGTTN TGTNTCAAC 200
CGCCTGCTTC TGTTTTAGCA ACGCAGTGTT TTGGTGGAAG TTGTGCCATG 250
TGTTCCACAA ANTCTTCCGA GATGGACACC CGAACGTCCT GAAGGACTTT 300
GTGAGATACA GAAATGAATT GAGTGACATG AGCAGGATGT GGGGCCACCT 350
GAGCGAGGGG TATGGCCAGC TGTGCAGCAT CTACCTGAAA CTGCTAAGAA 400
CCAAGATGGA GTACCACACC AAAAATCCCA GGTTCCTCAGG CAACCTGCAG 450
ATGAGTGACC GCCAGCTGGA CGAGGCTGGA GAAAGTGACG TGAACAACCT 500
TTTCCAGTTA ACAGTGAGGA TGTTTGACTA CCTGGAGTGT GAACTCAACC 550
TCTTCCAAAC AGTATTCAAC TCCCTGGACA TGTCCCGCTC TGTGTCCGTG 600
ACGGCAGCAG GGCAGTGCCG CCTCGCCCCG CTGATCCAGG TCATCTTGGA 650
CTGCAGCCAC CTTTATGACT ACACTGTCAA GCTTCTCTTC AAACCTCCACT 700
CCTGCCTCCC AGCTGACACC CTGCAAGGCC ACCGGGACCG CTTTCATGGAG 750

```

CAGTTTACAA	AGTTGAAAGA	TCTGTTCTAC	CGCTCCAGCA	ACCTGCAGTA	800
CTTCAAGCGG	CTCATTGAGA	TCCCCCAGCT	GCCTGAGAAC	CCACCCAACT	850
TCCTGCGAGC	CTCAGCCCTG	TCAGAACATA	TCAGCCCTGT	GGTGGTGATC	900
CCTGCAGAGG	CCTCATCCCC	CGACAGCGAG	CCAGTCCTAG	AGAAGGATGA	950
CCTCATGGAC	ATGGATGCCT	CTCAGCAGAA	TTTATTGAC	AACAAGTTTG	1000
ATGACATCTT	TGGCAGTTCA	TTCAGCAGTG	ATCCCTTCAA	TTTCAACAGT	1050
CAAAATGGTG	TGAACAAGGA	TGAGAAGGAC	CACTTAATTG	AGCGACTATA	1100
CAGAGAGATC	AGTGGATTGA	AGGCACAGCT	AGAAAACATG	AAGACTGAGA	1150
GCCAGCGGGT	TGTGCTGCAG	CTGAAGGGCC	ACGTCAGCGA	GCTGGAAGCA	1200
GATCTGGCCG	AGCAGCAGCA	CCTGCGGCAG	CAGGCGGCCG	ACGACTGTGA	1250
ATTCCTGCGG	GCAGAACTGG	ACGAGCTCAG	GAGGCAGCGG	GAGGACACCG	1300
AGAAGGCTCA	GCGGAGCCTG	TCTGAGATAG	AAAGGAAAGC	TCAAGCCAAT	1350
GAACAGCGAT	ATAGCAAGCT	AAAGGAGAAG	TACAGCGAGC	TGGTTCAGAA	1400
CCACGCTGAC	CTGCTGCGGA	AGAATGCAGA	GGTGACCAA	CAGGTGTCCA	1450
TGGCCAGACA	AGCCCAGGTA	GATTTGGAAC	GAGAGAAAAA	AGAGCTGGAG	1500
GATTCGTTGG	AGCGCATCAG	TGACCAGGGC	CAGCGGAAGA	CTCAAGAACA	1550
GCTGGAAGTT	CTAGAGAGCT	TGAAGCAGGA	ACTTGGCACA	AGCCAACGGG	1600
AGCTTCAGGT	TCTGCAAGGC	AGCCTGGAAG	CTTCTGCCCA	GTCAGAAGCA	1650
AACTGGGCAG	CCGAGTTCGC	CGAGCTAGAG	AAGGAGCGGG	ACAGCCTGGT	1700
GAGTGGCGCA	GCTCATAGGG	AGGAGGAATT	ATCTGCTCTT	CGGAAAGAAG	1750
TGCAGGACAC	TCAGCTCAAA	CTGGCCAGCA	CAGAGGAATC	TATGTGCCAG	1800
CTTGCCAAAG	ACCAACGAAA	AATGCTTCTG	GTGGGGTCCA	GGAAGGCTGC	1850
GGAGCAGGTG	ATACAAGACG	CCCTGAACCA	GCTTGAAGAA	CCTCCTCTCA	1900
TCAGCTGCGC	TGGGTCTGCA	GATCACCTCC	TCTCCACGGT	CACATCCATT	1950
TCCAGCTGCA	TCGAGCAACT	GGAGAAAAGC	TGGAGCCAGT	ATCTGGCCTG	2000
CCCAGAAGAC	ATCAGTGGAC	TTCTCCATTG	CATAACCCTG	CTGGCCCCACT	2050
TGACCAGCGA	CGCCATTGCT	CATGGTGCCA	CCACCTGCCT	CAGAGCCCCA	2100
CCTGAGCCTG	CCGACTCACT	GACCGAGGCC	TGTAAGCAGT	ATGGCAGGGA	2150
AACCCTCGCC	TACCTGGCCT	CCCTGGAGGA	AGAGGGAAGC	CTTGAGAATG	2200
CCGACAGCAC	AGCCATGAGG	AACTGCCTGA	GCAAGATCAA	GGCCATCGGC	2250
GAGGAGCTCC	TGCCCAGGGG	ACTGGACATC	AAGCAGGAGG	AGCTGGGGGA	2300
CCTGGTGGAC	AAGGAGATGG	CGGCCACTTC	AGCTGCTATT	GAAACTTGCA	2350
CGGCCAGAAT	AGAGGAGATG	CTCAGCAAAT	CCCGAGCAGG	AGACACAGGA	2400
GTCAAATTGG	AGGTGAATGA	AAGGATCCTT	CGTTGCTGTA	CCAGCCTCAT	2450
GCAAGCTATT	CAGGTGCTCA	TCGTGGCCTC	TAAGGACCTC	CAGAGAGAGA	2500
TTGTGGAGAG	CGGCAGGGGT	ACAGCATCCC	CTAAAGAGTT	TTATGCCAAG	2550
AACTCTCGAT	GGACAGAAGG	ACTTATCTCA	GCCTCCAAGG	CTGTGGGCTG	2600
GGGAGCCACT	GTCATGGTGG	ATGCAGCTGA	TCTGGTGGTA	CAAGGCAGAG	2650
GGAAATTTGA	GGAGCTAATG	GTGTGTTCTC	ATGAAATTGC	TGCTAGCACA	2700
GCCCAGCTTG	TGGCTGCATC	CAAGGTGAAA	GCTGATAAGG	ACAGCCCCAA	2750
CCTAGCCCAG	CTGCAGCAGG	CCTCTCGGGG	AGTGAACCAG	GCCACTGCCG	2800
GCGTTGTGGC	CTCAACCATT	TCCGGCAAAT	CACAGATCGA	AGAGACAGAC	2850
AACATGGACT	TCTCAAGCAT	GACGCTGACA	CAGATCAAAC	GCCAAGAGAT	2900
GGATTCTCAG	GTTAGGGTGC	TAGAGCTAGA	AAATGAATTG	CAGAAGGAGC	2950
GTCAAAAAC	GGGAGAGCTT	CGGAAAAAGC	ACTACGAGCT	TGCTGGTGTT	3000
GCTGAGGGCT	GGGAAGAAGG	AACAGAGGCA	TCTCCACCTA	CACTGCAAGA	3050
AGTGGTAACC	GAAAAAGAAT	AGAGCCAAAC	CAACACCCCA	TATGTCAGTG	3100
TAAATCCTTG	TTACCTATCT	CGTGTGTGTT	ATTTCCCCAG	CCACAGGCCA	3150
AATCCTTGGA	GTCCCAGGGG	CAGCCACACC	ACTGCCATTA	CCCAGTGCCG	3200
AGGACATGCA	TGACACTTCC	CAAAGATCCC	TCCATAGCGA	CACCCTTTCT	3250
GTTTGGAACC	ATGGTTCATCT	CTGTTCTTTT	CCCGCCTCCC	TAGTTAGCAT	3300

CCAGGCTGGC	CAGTGCTGCC	CATGAGCAAG	CCTAGGTACG	AAGAGGGGTG	3350
GTGGGGGGCA	GGGCCACTCA	ACAGAGAGGA	CCAACATCCA	GTCCTGCTGA	3400
CTATTTGACC	CCCACAACAA	TGGGTATCCT	TAATAGAGGA	GCTGCTTGTT	3450
GTTTGTTGAC	AGCTTGGA	GGGAAGATCT	TATGCCTTTT	CTTTTCTGTT	3500
TTCTTCTCAG	TCTTTTCAGT	TTCATCATTT	GCACAACTT	GTGAGCATCA	3550
GAGGGCTGAT	GGATTCCAAA	CCAGGACACT	ACCCTGAGAT	CTGCACAGTC	3600
AGAAGGACGG	CAGGAGTGTC	CTGGCTGTGA	ATGCCAAAGC	CATTCTCCCC	3650
CTCTTTGGGC	AGTGCCATGG	ATTTCCACTG	CTTCTTATGG	TGGTTGGTTG	3700
GGTTTTTTGG	TTTTGTTTTT	TTTTTTTAAG	TTTCACTCAC	ATAGCCAACT	3750
CTCCCAAAGG	GCACACCCCT	GGGGCTGAGT	CTCCAGGGCC	CCCCAACTGT	3800
GGTAGCTCCA	GCGATGGTGC	TGCCCAGGCC	TCTCGGTGCT	CCATCTCCGC	3850
CTCCACACTG	ACCAAGTGCT	GGCCACCCA	GTCCATGCTC	CAGGGTCAGG	3900
CGGAGCTGCT	GAGTGACAGC	TTTCCTCAA	AAGCAGAAGG	AGAGTGAGTG	3950
CCTTTCCCTC	CTAAAGCTGA	ATCCCGGCGG	AAAGCCTCTG	TCCGCCTTTA	4000
CAAGGGAGAA	GACAACAGAA	AGAGGGACAA	GAGGGTTCAC	ACAGCCCAGT	4050
TCCCGTGACG	AGGCTCAAAA	ACTTGATCAC	ATGCTTGAAT	GGAGCTGGTG	4100
AGATCAACAA	CACTACTTCC	CTGCCGGAAT	GAAGTGTCCG	TGAATGGTCT	4150
CTGTCAAGCG	GGCCGTCTCC	CTTGGCCCAG	AGACGGAGTG	TGGGAGTGAT	4200
TCCCAACTCC	TTTCTGCAGA	CGTCTGCCTT	GGCATCCTCT	TGAATAGGAA	4250
GATCGTTCCA	CTTTCTACGC	AATTGACAAA	CCCGGAAGAT	CAGATGCAAT	4300
TGCTCCCATC	AGGGAAGAAC	CCTATACTTG	GTTTGCTACC	CTTAGTATTT	4350
ATTACTAACC	TCCCTTAAGC	AGCAACAGCC	TACAAAGAGA	TGCTTGAGC	4400
AATCAGAACT	TCAGGTGTGA	CTCTAGCAAA	GCTCATCTTT	CTGCCC GGCT	4450
ACATCAGCCT	TCAAGAATCA	GAAGAAAGCC	AAGGTGCTGG	ACTGTTACTG	4500
ACTTGGATCC	CAAAGCAAGG	AGATCATTTG	GAGCTCTTGG	GTCAGAGAAA	4550
ATGAGAAAGG	ACAGAGCCAG	CGGCTCCAAC	TCCTTTCAGC	CACATGCCCC	4600
AGGCTCTCGC	TGCCCTGTGG	ACAGGATGAG	GACAGAGGGC	ACATGAACAG	4650
CTTGCCAGGG	ATGGGCAGCC	CAACAGCACT	TTTCCTCTTC	TAGATGGACC	4700
CCAGCATTTA	AGTGACCTTC	TGATCTTGGG	AAAACAGCGT	CTTCCTTCTT	4750
TATCTATAGC	AACTCATTTG	TGGTAGCCAT	CAAGCACTTC	GGAATT	4796

## (2) INFORMATION FOR SEQ ID NO: 4:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 924

(B) TYPE: protein

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: no

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(ix) FEATURE: Huntingtin-interacting protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Met	Ser	Arg	Met	Trp	Gly	His	Leu	Ser	Glu	Gly	Tyr	Gly	Gln	Leu
1				5				10						15

Cys	Ser	Ile	Tyr	Leu	Lys	Leu	Leu	Arg	Thr	Lys	Met	Glu	Tyr	His
				20				25						30

Thr	Lys	Asn	Pro	Arg	Phe	Pro	Gly	Asn	Leu	Gln	Met	Ser	Asp	Arg	
				35					40					45	
Gln	Leu	Asp	Glu	Ala	Gly	Glu	Ser	Asp	Val	Asn	Asn	Phe	Phe	Gln	
				50					55					60	
Leu	Thr	Val	Glu	Met	Phe	Asp	Tyr	Leu	Glu	Cys	Glu	Leu	Asn	Leu	
				65					70					75	
Phe	Gln	Thr	Val	Phe	Asn	Ser	Leu	Asp	Met	Ser	Arg	Ser	Val	Ser	
				80					85					90	
Val	Thr	Ala	Ala	Gly	Gln	Cys	Arg	Leu	Ala	Pro	Leu	Ile	Gln	Val	
				95					100					105	
Ile	Leu	Asp	Cys	Ser	His	Leu	Tyr	Asp	Tyr	Thr	Val	Lys	Leu	Leu	
				110					115					120	
Phe	Lys	Leu	His	Ser	Cys	Leu	Pro	Ala	Asp	Thr	Leu	Gln	Gly	His	
				125					130					135	
Arg	Asp	Arg	Phe	Met	Glu	Gln	Phe	Thr	Lys	Leu	Lys	Asp	Leu	Phe	
				140					145					150	
Tyr	Arg	Ser	Ser	Asn	Leu	Gln	Tyr	Phe	Lys	Arg	Leu	Ile	Gln	Ile	
				155					160					165	
Pro	Gln	Leu	Pro	Glu	Asn	Pro	Pro	Asn	Phe	Leu	Arg	Ala	Ser	Ala	
				170					175					180	
Leu	Ser	Glu	His	Ile	Ser	Pro	Val	Val	Val	Ile	Pro	Ala	Glu	Ala	
				185					190					195	
Ser	Ser	Pro	Asp	Ser	Glu	Pro	Val	Leu	Glu	Lys	Asp	Asp	Leu	Met	
				200					205					210	
Asp	Met	Asp	Ala	Ser	Gln	Gln	Asn	Leu	Phe	Asp	Asn	Lys	Phe	Asp	
				215					220					225	
Asp	Ile	Phe	Gly	Ser	Ser	Phe	Ser	Ser	Asp	Pro	Phe	Asn	Phe	Asn	
				230					235					240	
Ser	Gln	Asn	Gly	Val	Asn	Lys	Asp	Glu	Lys	Asp	His	Leu	Ile	Glu	
				245					250					255	
Arg	Leu	Tyr	Arg	Glu	Ile	Ser	Gly	Leu	Lys	Ala	Gln	Leu	Glu	Asn	
				260					265					270	
Met	Lys	Thr	Glu	Ser	Gln	Arg	Val	Val	Leu	Gln	Leu	Lys	Gly	His	
				275					280					285	

Val	Ser	Glu	Leu	Glu	Ala	Asp	Leu	Ala	Glu	Gln	Gln	His	Leu	Arg	290	295	300
Gln	Gln	Ala	Ala	Asp	Asp	Cys	Glu	Phe	Leu	Arg	Ala	Glu	Leu	Asp	305	310	315
Glu	Leu	Arg	Arg	Gln	Arg	Glu	Asp	Thr	Glu	Lys	Ala	Gln	Arg	Ser	320	325	330
Leu	Ser	Glu	Ile	Glu	Arg	Lys	Ala	Gln	Ala	Asn	Glu	Gln	Arg	Tyr	335	340	345
Ser	Lys	Leu	Lys	Glu	Lys	Tyr	Ser	Glu	Leu	Val	Gln	Asn	His	Ala	350	355	360
Asp	Leu	Leu	Arg	Lys	Asn	Ala	Glu	Val	Thr	Lys	Gln	Val	Ser	Met	365	370	375
Ala	Arg	Gln	Ala	Gln	Val	Asp	Leu	Glu	Arg	Glu	Lys	Lys	Glu	Leu	380	385	390
Glu	Asp	Ser	Leu	Glu	Arg	Ile	Ser	Asp	Gln	Gly	Gln	Arg	Lys	Thr	395	400	405
Gln	Glu	Gln	Leu	Glu	Val	Leu	Glu	Ser	Leu	Lys	Gln	Glu	Leu	Gly	410	415	420
Thr	Ser	Gln	Arg	Glu	Leu	Gln	Val	Leu	Gln	Gly	Ser	Leu	Glu	Thr	425	430	435
Ser	Ala	Gln	Ser	Glu	Ala	Asn	Trp	Ala	Ala	Glu	Phe	Ala	Glu	Leu	440	445	450
Glu	Lys	Glu	Arg	Asp	Ser	Leu	Val	Ser	Gly	Ala	Ala	His	Arg	Glu	455	460	465
Glu	Glu	Leu	Ser	Ala	Leu	Arg	Lys	Glu	Leu	Gln	Asp	Thr	Gln	Leu	470	475	480
Lys	Leu	Ala	Ser	Thr	Glu	Glu	Ser	Met	Cys	Gln	Leu	Ala	Lys	Asp	485	490	495
Gln	Arg	Lys	Met	Leu	Leu	Val	Gly	Ser	Arg	Lys	Ala	Ala	Glu	Gln	500	505	510
Val	Ile	Gln	Asp	Ala	Leu	Asn	Gln	Leu	Glu	Glu	Pro	Pro	Leu	Ile	515	520	525
Ser	Cys	Ala	Gly	Ser	Ala	Asp	His	Leu	Leu	Ser	Thr	Val	Thr	Ser	530	535	540



Ile Ser Ser Cys	Ile Glu Gln Leu Glu Lys Ser Trp Ser Gln Tyr	545	550	555
Leu Ala Cys Pro	Glu Asp Ile Ser Gly Leu Leu His Ser Ile Thr	560	565	570
Leu Leu Ala His	Leu Thr Ser Asp Ala Ile Ala His Gly Ala Thr	575	580	585
Thr Cys Leu Arg	Ala Pro Pro Glu Pro Ala Asp Ser Leu Thr Glu	590	595	600
Ala Cys Lys Gln	Tyr Gly Arg Glu Thr Leu Ala Tyr Leu Ala Ser	605	610	615
Leu Glu Glu Glu	Gly Ser Leu Glu Asn Ala Asp Ser Thr Ala Met	620	625	630
Arg Asn Cys Leu	Ser Lys Ile Lys Ala Ile Gly Glu Glu Leu Leu	635	640	645
Pro Arg Gly Leu	Asp Ile Lys Gln Glu Glu Leu Gly Asp Leu Val	650	655	660
Asp Lys Glu Met	Ala Ala Thr Ser Ala Ala Ile Glu Thr Cys Thr	665	670	675
Ala Arg Ile Glu	Glu Met Leu Ser Lys Ser Arg Ala Gly Asp Thr	680	685	690
Gly Val Lys Leu	Glu Val Asn Glu Arg Ile Leu Arg Cys Cys Thr	695	700	705
Ser Leu Met Gln	Ala Ile Gln Val Leu Ile Val Ala Ser Lys Asp	710	715	720
Leu Gln Arg Glu	Ile Val Glu Ser Gly Arg Gly Thr Ala Ser Pro	725	730	735
Lys Glu Phe Tyr	Ala Lys Asn Ser Arg Trp Thr Glu Gly Leu Ile	740	745	750
Ser Ala Ser Lys	Ala Val Gly Trp Gly Ala Thr Val Met Val Asp	765	770	775
Ala Ala Asp Leu	Val Val Gln Gly Arg Gly Lys Phe Glu Glu Leu	780	785	790
Met Val Cys Ser	His Glu Ile Ala Ala Ser Thr Ala Gln Leu Val	795	800	805

Ala	Ala	Ser	Lys	Val	Lys	Ala	Asp	Lys	Asp	Ser	Pro	Asn	Leu	Ala	
				810					815					820	
Gln	Leu	Gln	Gln	Ala	Ser	Arg	Gly	Val	Asn	Gln	Ala	Thr	Ala	Gly	
				825					830					835	
Val	Val	Ala	Ser	Thr	Ile	Ser	Gly	Lys	Ser	Gln	Ile	Glu	Glu	Thr	
				840					845					850	
Asp	Asn	Met	Asp	Phe	Ser	Ser	Met	Thr	Leu	Thr	Gln	Ile	Lys	Arg	
				855					860					865	
Gln	Glu	Met	Asp	Ser	Gln	Val	Arg	Val	Leu	Glu	Leu	Glu	Asn	Glu	
				870					875					880	
Leu	Gln	Lys	Glu	Arg	Gln	Lys	Leu	Gly	Glu	Leu	Arg	Lys	Lys	His	
				885					890					895	
Tyr	Glu	Leu	Ala	Gly	Val	Ala	Glu	Gly	Trp	Glu	Glu	Gly	Thr	Glu	
				900					905					910	
Ala	Ser	Pro	Pro	Thr	Leu	Gln	Glu	Val	Val	Thr	Glu	Lys	Glu		
				915					920				924		

## (2) INFORMATION FOR SEQ ID NO: 5

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1090

(B) TYPE: protein

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: no

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(ix) FEATURE: Huntingtin-interacting protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

Met	Leu	Leu	Cys	Gln	Gly	Ser	Glu	Trp	Arg	Arg	Asp	Gln	Gln	Leu	
				5					10					15	
Gly	Thr	Ala	Asn	Ala	Arg	Gln	Trp	Cys	Pro	Leu	Pro	Gln	Asp	Ala	
				20					25					30	
Gln	Pro	Ala	Gly	Ser	Trp	Glu	Arg	Cys	Pro	Pro	Leu	Pro	Pro	Ala	
				35					40					45	
Gly	Arg	Leu	Gln	Gly	Thr	Asp	His	Pro	Trp	Gly	Trp	Gly	Arg	Leu	
				50					55					60	

Ala	Gly	Gly	Gly	Glu	Arg	Gly	Gly	Leu	Trp	Glu	Gly	Leu	Ser	His
				65					70					75
Ser	Gln	Arg	Leu	Ile	His	Leu	Ile	Leu	Leu	Ser	Leu	Pro	Leu	Leu
				80					85					90
Val	Phe	Gln	Thr	Val	Ser	Ile	Asn	Lys	Ala	Ile	Asn	Thr	Gln	Glu
				95					100					105
Val	Ala	Val	Lys	Glu	Lys	His	Ala	Arg	Thr	Cys	Ile	Leu	Gly	Thr
				110					115					120
His	His	Glu	Lys	Gly	Ala	Gln	Thr	Phe	Trp	Ser	Val	Val	Asn	Arg
				125					130					135
Leu	Pro	Leu	Ser	Ser	Asn	Ala	Val	Leu	Cys	Trp	Lys	Phe	Cys	His
				140					145					150
Val	Phe	His	Lys	Leu	Leu	Arg	Asp	Gly	His	Pro	Asn	Val	Leu	Lys
				155					160					165
Asp	Ser	Leu	Arg	Tyr	Arg	Asn	Glu	Leu	Ser	Asp	Met	Ser	Arg	Met
				170					175					180
Trp	Gly	His	Leu	Ser	Glu	Gly	Tyr	Gly	Gln	Leu	Cys	Ser	Ile	Tyr
				185					190					195
Leu	Lys	Leu	Leu	Arg	Thr	Lys	Met	Glu	Tyr	His	Thr	Lys	Asn	Pro
				200					205					210
Arg	Phe	Pro	Gly	Asn	Leu	Gln	Met	Ser	Asp	Arg	Gln	Leu	Asp	Glu
				215					220					225
Ala	Gly	Glu	Ser	Asp	Val	Asn	Asn	Phe	Phe	Gln	Leu	Thr	Val	Glu
				230					235					240
Met	Phe	Asp	Tyr	Leu	Glu	Cys	Glu	Leu	Asn	Leu	Phe	Gln	Thr	Val
				245					250					255
Phe	Asn	Ser	Leu	Asp	Met	Ser	Arg	Ser	Val	Ser	Val	Thr	Ala	Ala
				260					265					270
Gly	Gln	Cys	Arg	Leu	Ala	Pro	Leu	Ile	Gln	Val	Ile	Leu	Asp	Cys
				275					288					285
Ser	His	Leu	Tyr	Asp	Tyr	Thr	Val	Lys	Leu	Leu	Phe	Lys	Leu	His
				290					295					300
Ser	Cys	Leu	Pro	Ala	Asp	Thr	Leu	Gln	Gly	His	Arg	Asp	Arg	Phe
				305					310					315

Met	Glu	Gln	Phe	Thr	Lys	Leu	Lys	Asp	Leu	Phe	Tyr	Arg	Ser	Ser	
				320					325					330	
Asn	Leu	Gln	Tyr	Phe	Lys	Arg	Leu	Ile	Gln	Ile	Pro	Gln	Leu	Pro	
				335					340					345	
Glu	Asn	Pro	Pro	Asn	Phe	Leu	Arg	Ala	Ser	Ala	Leu	Ser	Glu	His	
				350					355					360	
Ile	Ser	Pro	Val	Val	Val	Ile	Pro	Ala	Glu	Ala	Ser	Ser	Pro	Asp	
				365					370					375	
Ser	Glu	Pro	Val	Leu	Glu	Lys	Asp	Asp	Leu	Met	Asp	Met	Asp	Ala	
				380					385					390	
Ser	Gln	Gln	Asn	Leu	Phe	Asp	Asn	Lys	Phe	Asp	Asp	Ile	Phe	Gly	
				395					400					405	
Ser	Ser	Phe	Ser	Ser	Asp	Pro	Phe	Asn	Phe	Asn	Ser	Gln	Asn	Gly	
				410					415					420	
Val	Asn	Lys	Asp	Glu	Lys	Asp	His	Leu	Ile	Glu	Arg	Leu	Tyr	Arg	
				425					430					435	
Glu	Ile	Ser	Gly	Leu	Lys	Ala	Gln	Leu	Glu	Asn	Met	Lys	Thr	Glu	
				440					445					450	
Ser	Gln	Arg	Val	Val	Leu	Gln	Leu	Lys	Gly	His	Val	Ser	Glu	Leu	
				455					460					465	
Glu	Ala	Asp	Leu	Ala	Glu	Gln	Gln	His	Leu	Arg	Gln	Gln	Ala	Ala	
				470					475					480	
Asp	Asp	Cys	Glu	Phe	Leu	Arg	Ala	Glu	Leu	Asp	Glu	Leu	Arg	Arg	
				485					490					495	
Gln	Arg	Glu	Asp	Thr	Glu	Lys	Ala	Gln	Arg	Ser	Leu	Ser	Glu	Ile	
				500					505					510	
Glu	Arg	Lys	Ala	Gln	Ala	Asn	Glu	Gln	Arg	Tyr	Ser	Lys	Leu	Lys	
				515					520					525	
Glu	Lys	Tyr	Ser	Glu	Leu	Val	Gln	Asn	His	Ala	Asp	Leu	Leu	Arg	
				530					535					540	
Lys	Asn	Ala	Glu	Val	Thr	Lys	Gln	Val	Ser	Met	Ala	Arg	Gln	Ala	
				545					550					555	
Gln	Val	Asp	Leu	Glu	Arg	Glu	Lys	Lys	Glu	Leu	Glu	Asp	Ser	Leu	
				560					565					570	

Glu	Arg	Ile	Ser	Asp	Gln	Gly	Gln	Arg	Lys	Thr	Gln	Glu	Gln	Leu
				575					588					585
Glu	Val	Leu	Glu	Ser	Leu	Lys	Gln	Glu	Leu	Ala	Thr	Ser	Gln	Arg
				590					595					600
Glu	Leu	Gln	Val	Leu	Gln	Gly	Ser	Leu	Glu	Thr	Ser	Ala	Gln	Ser
				605					610					615
Glu	Ala	Asn	Trp	Ala	Ala	Glu	Phe	Ala	Glu	Leu	Glu	Lys	Glu	Arg
				620					625					630
Asp	Ser	Leu	Val	Ser	Gly	Ala	Ala	His	Arg	Glu	Glu	Glu	Leu	Ser
				635					640					645
Ala	Leu	Arg	Lys	Glu	Leu	Gln	Asp	Thr	Gln	Leu	Lys	Leu	Ala	Ser
				650					655					660
Thr	Glu	Glu	Ser	Met	Cys	Gln	Leu	Ala	Lys	Asp	Gln	Arg	Lys	Met
				665					670					675
Leu	Leu	Val	Gly	Ser	Arg	Lys	Ala	Ala	Glu	Gln	Val	Ile	Gln	Asp
				680					685					690
Ala	Leu	Asn	Gln	Leu	Glu	Glu	Pro	Pro	Leu	Ile	Ser	Cys	Ala	Gly
				695					700					705
Ser	Ala	Asp	His	Leu	Leu	Ser	Thr	Val	Thr	Ser	Ile	Ser	Ser	Cys
				710					715					720
Ile	Glu	Gln	Leu	Glu	Lys	Ser	Trp	Ser	Gln	Tyr	Leu	Ala	Cys	Pro
				725					730					735
Glu	Asp	Ile	Ser	Gly	Leu	Leu	His	Ser	Ile	Thr	Leu	Leu	Ala	His
				740					745					750
Leu	Thr	Ser	Asp	Ala	Ile	Ala	His	Gly	Ala	Thr	Thr	Cys	Leu	Arg
				755					760					765
Ala	Pro	Pro	Glu	Pro	Ala	Asp	Ser	Leu	Thr	Glu	Ala	Cys	Lys	Gln
				770					775					780
Tyr	Gly	Arg	Glu	Thr	Leu	Ala	Tyr	Leu	Ala	Ser	Leu	Glu	Glu	Glu
				785					790					795
Gly	Ser	Leu	Glu	Asn	Ala	Asp	Ser	Thr	Ala	Met	Arg	Asn	Cys	Leu
				800					805					810
Ser	Lys	Ile	Lys	Ala	Ile	Gly	Glu	Glu	Leu	Leu	Pro	Arg	Gly	Leu
				815					820					825

Asp	Ile	Lys	Gln	Glu	Glu	Leu	Gly	Asp	Leu	Val	Asp	Lys	Glu	Met
				830					835					840
Ala	Ala	Thr	Ser	Ala	Ala	Ile	Glu	Thr	Ala	Thr	Ala	Arg	Ile	Glu
				845					850					855
Glu	Met	Leu	Ser	Lys	Ser	Arg	Ala	Gly	Asp	Thr	Gly	Val	Lys	Leu
				860					865					870
Glu	Val	Asn	Glu	Arg	Ile	Leu	Gly	Cys	Cys	Thr	Ser	Leu	Met	Gln
				875					888					885
Ala	Ile	Gln	Val	Leu	Ile	Val	Ala	Ser	Lys	Asp	Leu	Gln	Arg	Glu
				890					895					900
Ile	Val	Glu	Ser	Gly	Arg	Gly	Thr	Ala	Ser	Pro	Lys	Glu	Phe	Tyr
				905					910					915
Ala	Lys	Asn	Ser	Arg	Trp	Thr	Glu	Gly	Leu	Ile	Ser	Ala	Ser	Lys
				920					925					930
Ala	Val	Gly	Trp	Gly	Ala	Thr	Val	Met	Val	Asp	Ala	Ala	Asp	Leu
				935					940					945
Val	Val	Gln	Gly	Arg	Gly	Lys	Phe	Glu	Glu	Leu	Met	Val	Cys	Ser
				950					955					960
His	Glu	Ile	Ala	Ala	Ser	Thr	Ala	Gln	Leu	Val	Ala	Ala	Ser	Lys
				965					970					975
Val	Lys	Ala	Asp	Lys	Asp	Ser	Pro	Asn	Leu	Ala	Gln	Leu	Gln	Gln
				980					985					990
Ala	Ser	Arg	Gly	Val	Asn	Gln	Ala	Thr	Ala	Gly	Val	Val	Ala	Ser
				995					1000					1005
Thr	Ile	Ser	Gly	Lys	Ser	Gln	Ile	Glu	Glu	Thr	Asp	Asn	Met	Asp
				1010					1015					1020
Phe	Ser	Ser	Met	Thr	Leu	Thr	Gln	Ile	Lys	Arg	Gln	Glu	Met	Asp
				1025					1030					1035
Ser	Gln	Val	Arg	Val	Leu	Glu	Leu	Glu	Asn	Glu	Leu	Gln	Lys	Glu
				1040					1045					1050
Arg	Gln	Lys	Leu	Gly	Glu	Leu	Arg	Lys	Lys	His	Tyr	Glu	Leu	Ala
				1055					1060					1065
Gly	Val	Ala	Glu	Gly	Trp	Glu	Glu	Gly	Thr	Glu	Ala	Ser	Pro	Pro
				1070					1075					1080

Thr Leu Gln Glu Val Val Thr Glu Lys Glu  
1085 1090

## (2) INFORMATION FOR SEQ ID NO:6:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3301

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(ix) FEATURE: cDNA for Huntingtin-interacting protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

CGGTGAGCTG	GAGGAGCAGC	GGAAGCAGAA	GCAGAAGGCC	CTGGTGGATA	50
ATGAGCAGCT	CCGCCACGAG	CTGGCCCAGC	TGAGGGCTGC	CCAGCTGGAG	100
CGCGAGCGGA	GCCAGGGCCT	GCGTGAGGAG	GCTGAGAGGA	AGGCCAGTGC	150
CACGGAGGCG	CGCTACAACA	AGCTGAAGGA	AAAGCACAGT	GAGCTCGTCC	200
ATGTGCACGC	GGAGCTGCTC	AGAAAGAACG	CGGACACAGC	CAAGCAGCTG	250
ACGGTGACGC	AGCAAAGCCA	GGAGGAGGTG	GCGCGGGTGA	AGGAGCAGCT	300
GGCCTTCCAG	GTGGAGCAGG	TGAAGCGGGA	GTCGGAGTTG	AAGCTAGAGG	350
AGAAGAGCGA	CCAGCAGGAG	AAGCTCAAGA	GGGAGCTGGA	GGCCAAGGCC	400
GGAGAGCTGG	CCCGCGCGCA	GGAGGCCCTG	AGCCACACAG	AGCAGAGCAA	450
GTCGGAGCTG	AGCTCACGGC	TGGACACACT	GAGTGCGGAG	AAGGATGCTC	500
TGAGTGGAGC	TGTGCGGCAG	CGGGAGGCAG	ACCTGCTGGC	GGCGCAGAGC	550
CTGGTGCGCG	AGACAGAGGC	GGCGCTGAGC	CGGGAGCAGC	AGCGCAGCTC	600
CCAGGAGCAG	GGCGAGTTGC	AGGGCCGGCT	GGCAGAGAGG	GAGTCTCAGG	650
AGCAGGGGCT	GCGGCAGAGG	CTGCTGGACG	AGCAGTTCGC	AGTGTTGCGG	700
GGCGCTGCTG	CCGAGGCCGC	GGGCATCCTG	CAGGATGCCG	TGAGCAAGCT	750
GGACGACCCC	CTGCACCTGC	GCTGTACCAG	CTCCCCAGAC	TACCTGGTGA	800
GCAGGGCCCA	GGAGGCCTTG	GATGCCGTGA	GCACCCTGGA	GGAGGGCCAC	850
GCCCAGTACC	TGACCTCCTT	GGCAGACGCC	TCCGCCCTGG	TGGCAGCTCT	900
GACCCGCTTC	TCCCACCTGG	CTGCGGATAC	CATCATCAAT	GGCGGTGCCA	950
CCTCGCACCT	GGCTCCCACC	GACCCTGCCG	ACCGCCTCAT	AGACACCTGC	1000
AGGGAGTGCG	GGGCCCGGGC	TCTGGAGCTC	ATGGGGCAGC	TGCAGGACCA	1050
GCAGGCTCTG	CGGCACATGC	AGGCCAGCCT	GGTGCGGACA	CCCCTGCAGG	1100
GCATCCTTCA	GCTGGGCCAA	GAAGTGA AAC	CCAAGAGCCT	AGATGTGCGG	1150
CAGGAGGAGC	TGGGGGCCGT	GGTCGACAAG	GAGATGGCGG	CCACATCCGC	1200
AGCCATTGAA	GATGCTGTGC	GGAGGATTGA	GGACATGATG	AACCAGGCAC	1250
GCCACGCCAG	CTCGGGGGTG	AAGCTGGAGG	TGAACGAGAG	GATCCTCAAC	1300
TCCTGCACAG	ACCTGATGAA	GGCTATCCGG	CTCCTGGTGA	CGACATCCAC	1350
TAGCCTGCAG	AAGGAGATCG	TGGAGAGCGG	CAGGGGGGCA	GCCACGCAGC	1400
AGGAATTTTA	CGCCAAGAAC	TCGCGCTGGA	CCGAAGGCCT	CATCTCGGCC	1450
TCCAAGGCTG	TGGGCTGGGG	AGCCACACAG	CTGGTGGAGG	CAGCTGACAA	1500
GGTGGTGCTT	CACACGGGCA	AGTATGAGGA	GCTCATCGTC	TGCTCCCACG	1550
AGATCGCAGC	CAGCACGGCC	CAGCTGGTGG	CGGCCTCCAA	GGTGAAGGCC	1600

AACAAAGCACA GCCCCACCT GAGCCGCCTG CAGGAATGTT CTCGCACAGT 1650  
 CAATGAGAGG GCTGCCAATG TGGTGGCCTC CACCAAGTCA GGCCAGGAGC 1700  
 AGATTGAGGA CAGAGACACC ATGGATTCTT CCGGCCTGTC CCTCATCAAG 1750  
 CTGAAGAAGC AGGAGATGGA GACGCAGGTG CGTGTCTTGG AGCTGGAGAA 1800  
 GACGCTGGAG GCTGAACGCA TGC GGCTGGG GGAGTTGCGG AAGCAACACT 1850  
 ACGTGCTGGC TGGGGCATCA GGCAGCCCTG GAGAGGAGGT GGCCATCCGG 1900  
 CCCAGCACTG CCCCCGAAG TGTAACCACC AAGAAACCAC CCCTGGCCCA 1950  
 GAAGCCCAGC GTGGCCCCCA GACAGGACCA CCAGCTTGAC AAAAAGGATG 2000  
 GCATCTACCC AGCTCAACTC GTGAACACT AGGCCCCCA GGGGTCCAGC 2050  
 AGGGTGGCTG GTGACAGGCC TGGGCCTCTG CAACTGCCCT GACAGGACCG 2100  
 AGAGGCCTTG CCCCTCCACC TGGTGCCCAA GCCTCCCGCC CCACCGTCTG 2150  
 GATCAATGTC CTCAAGGCCC CTGGCCCTTA CTGAGCCTGC AGGGTCCTGG 2200  
 GCCATGTGGG TGGTGCTTCT GGATGTGAGT CTCTTATTTA TCTGCAGAAG 2250  
 GAACTTTGGG GTGCAGCCAG GACCCGCTAG GCCTGAGCCT CAACTCTTCA 2300  
 GAAAATAGTG TTTTAAATAT TCCTCTTCAG AAAATAGTGT TTTTAATATT 2350  
 CCGAGCTAGA GCTCTTCTTC CTACGTTTGT AGTCAGCACA CTGGGAAACC 2400  
 GGGCCAGCGT GGGGCTCCCT GCCTTCTGGA CTCCTGAAGG TCGTGGATGG 2450  
 ATGGAAGGCA CACAGCCCGT GCCGGCTGAT GGGACGAGGG TCAGGCATCC 2500  
 TGTCTGTGGC CTTCTGGGGC ACCGATTCTA CCAGGCCCTC CAGCTGCGTG 2550  
 GTCTCCGCAG ACCAGGCTCT GTGTGGGCTA GAGGAATGTC GCCCATTACC 2600  
 TCCTCAGGCC CTGGCCCTCG GGCCTCCGTG ATGGGAGCCC CCCAGGAGGG 2700  
 GTCAGATGCT GGAAGGGGCC GCTTTCTGGG GAGTGAGGTG AGACATAGCG 2750  
 GCCCAGGCGC TGCCTTCACT CCTGGAGTTT CCATTTCCAG CTGGAATCTG 2800  
 CAGCCACCCC CATTTCTGT TTTCCATTCC CCCGTTCTGG CCGCGCCCCA 2850  
 CTGCCCACCT GAAGGGGTGG TTTCCAGCCC TCCGGAGAGT GGGCTTGGCC 2900  
 CTAGGCCCTC CAGCTCAGCC AGAAAAAGCC CAGAAACCCA GGTGCTGGAC 2950  
 CAGGGCCCTC AGGGAGGGAC CCTGCGGCTA GAGTGGGCTA GGCCCTGGCT 3000  
 TTGCCCCGTC GATTTGAACG AATGTGTGTC CCTTGAGCCC AAGGAGAGCG 3050  
 GCAGGAGGGG TGGGACCAGG CTGGGAGGAC AGAGCCAGCA GCTGCCATGC 3100  
 CCTCCTGCTC CCCCCACCCC AGCCCTAGCC CTTTAGCCTT TCACCCTGTG 3150  
 CTCTGGAAAG GCTACCAAAT ACTGGCCAAG GTCAGGAGGA GCAAAAATGA 3200  
 GCCAGCACCA GCGCCTTGGC TTTGTGTTAG CATTTCTTCC TGAAGTGTTT 3250  
 TGTGGAAT AAAATGCACT TTGACTGTTA AAAAAAAAAA AAAAAAAAAA 3300  
 A 3301

## (2) INFORMATION FOR SEQ ID NO: 7

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 676

(B) TYPE: protein

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: no

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(ix) FEATURE: Huntingtin-interacting protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

Gly Glu Leu Glu Glu Gln Arg Lys Gln Lys Gln Lys Ala Leu Val

5

10

15



Asp	Asn	Glu	Gln	Leu	Arg	His	Glu	Leu	Ala	Gln	Leu	Arg	Ala	Ala	
				20					25					30	
Gln	Leu	Glu	Arg	Glu	Arg	Ser	Gln	Gly	Leu	Arg	Glu	Glu	Ala	Glu	
				35					40					45	
Arg	Lys	Ala	Ser	Ala	Thr	Glu	Ala	Arg	Tyr	Asn	Lys	Leu	Lys	Glu	
				50					55					60	
Lys	His	Ser	Glu	Leu	Val	His	Val	His	Ala	Glu	Leu	Leu	Arg	Lys	
				65					70					75	
Asn	Ala	Asp	Thr	Ala	Lys	Gln	Leu	Thr	Val	Thr	Gln	Gln	Ser	Gln	
				80					85					90	
Glu	Glu	Val	Ala	Arg	Val	Lys	Glu	Gln	Leu	Ala	Phe	Gln	Val	Glu	
				95					100					105	
Gln	Val	Lys	Arg	Glu	Ser	Glu	Leu	Lys	Leu	Glu	Glu	Lys	Ser	Asp	
				110					115					120	
Gln	Gln	Glu	Lys	Leu	Lys	Arg	Glu	Leu	Glu	Ala	Lys	Ala	Gly	Glu	
				125					130					135	
Leu	Ala	Arg	Ala	Gln	Glu	Ala	Leu	Ser	His	Thr	Glu	Gln	Ser	Lys	
				140					145					150	
Ser	Glu	Leu	Ser	Ser	Arg	Leu	Asp	Thr	Leu	Ser	Ala	Glu	Lys	Asp	
				155					160					165	
Ala	Leu	Ser	Gly	Ala	Val	Arg	Gln	Arg	Glu	Ala	Asp	Leu	Leu	Ala	
				170					175					180	
Ala	Gln	Ser	Leu	Val	Arg	Glu	Thr	Glu	Ala	Ala	Leu	Ser	Arg	Glu	
				185					190					195	
Gln	Gln	Arg	Ser	Ser	Gln	Glu	Gln	Gly	Glu	Leu	Gln	Gly	Arg	Leu	
				200					205					210	
Ala	Glu	Arg	Glu	Ser	Gln	Glu	Gln	Gly	Leu	Arg	Gln	Arg	Leu	Leu	
				215					220					225	
Asp	Glu	Gln	Phe	Ala	Val	Leu	Arg	Gly	Ala	Ala	Ala	Glu	Ala	Ala	
				230					235					240	
Gly	Ile	Leu	Gln	Asp	Ala	Val	Ser	Lys	Leu	Asp	Asp	Pro	Leu	His	
				245					250					255	
Leu	Arg	Cys	Thr	Ser	Ser	Pro	Asp	Tyr	Leu	Val	Ser	Arg	Ala	Gln	
				260					265					270	

Glu	Ala	Leu	Asp	Ala	Val	Ser	Thr	Leu	Glu	Glu	Gly	His	Ala	Gln	
				275					288					285	
Tyr	Leu	Thr	Ser	Leu	Ala	Asp	Ala	Ser	Ala	Leu	Val	Ala	Ala	Leu	
				290					295					300	
Thr	Arg	Phe	Ser	His	Leu	Ala	Ala	Asp	Thr	Ile	Ile	Asn	Gly	Gly	
				305					310					315	
Ala	Thr	Ser	His	Leu	Ala	Pro	Thr	Asp	Pro	Ala	Asp	Arg	Leu	Ile	
				320					325					330	
Asp	Thr	Cys	Arg	Glu	Cys	Gly	Ala	Arg	Ala	Leu	Glu	Leu	Met	Gly	
				335					340					345	
Gln	Leu	Gln	Asp	Gln	Gln	Ala	Leu	Arg	His	Met	Gln	Ala	Ser	Leu	
				350					355					360	
Val	Arg	Thr	Pro	Leu	Gln	Gly	Ile	Leu	Gln	Leu	Gly	Gln	Glu	Leu	
				365					370					375	
Lys	Pro	Lys	Ser	Leu	Asp	Val	Arg	Gln	Glu	Glu	Leu	Gly	Ala	Val	
				380					385					390	
Val	Asp	Lys	Glu	Met	Ala	Ala	Thr	Ser	Ala	Ala	Ile	Glu	Asp	Ala	
				395					400					405	
Val	Arg	Arg	Ile	Glu	Asp	Met	Met	Asn	Gln	Ala	Arg	His	Ala	Ser	
				410					415					420	
Ser	Gly	Val	Lys	Leu	Glu	Val	Asn	Glu	Arg	Ile	Leu	Asn	Ser	Cys	
				425					430					435	
Thr	Asp	Leu	Met	Lys	Ala	Ile	Arg	Leu	Leu	Val	Thr	Thr	Ser	Thr	
				440					445					450	
Ser	Leu	Gln	Lys	Glu	Ile	Val	Glu	Ser	Gly	Arg	Gly	Ala	Ala	Thr	
				455					460					465	
Gln	Gln	Glu	Phe	Tyr	Ala	Lys	Asn	Ser	Arg	Trp	Thr	Glu	Gly	Leu	
				470					475					480	
Ile	Ser	Ala	Ser	Lys	Ala	Val	Gly	Trp	Gly	Ala	Thr	Gln	Leu	Val	
				485					490					495	
Glu	Ala	Ala	Asp	Lys	Val	Val	Leu	His	Thr	Gly	Lys	Tyr	Glu	Glu	
				500					505					510	
Leu	Ile	Val	Cys	Ser	His	Glu	Ile	Ala	Ala	Ser	Thr	Ala	Gln	Leu	
				515					520					525	

Val	Ala	Ala	Ser	Lys	Val	Lys	Ala	Asn	Lys	His	Ser	Pro	His	Leu
				530					535					540
Ser	Arg	Leu	Gln	Glu	Cys	Ser	Arg	Thr	Val	Asn	Glu	Arg	Ala	Ala
				545					550					555
Asn	Val	Val	Ala	Ser	Thr	Lys	Ser	Gly	Gln	Glu	Gln	Ile	Glu	Asp
				560					565					570
Arg	Asp	Thr	Met	Asp	Phe	Ser	Gly	Leu	Ser	Leu	Ile	Lys	Leu	Lys
				575					588					585
Lys	Gln	Glu	Met	Glu	Thr	Gln	Val	Arg	Val	Leu	Glu	Leu	Glu	Lys
				590					595					600
Thr	Leu	Glu	Ala	Glu	Arg	Met	Arg	Leu	Gly	Glu	Leu	Arg	Lys	Gln
				605					610					615
His	Tyr	Val	Leu	Ala	Gly	Ala	Ser	Gly	Ser	Pro	Gly	Glu	Glu	Val
				620					625					630
Ala	Ile	Arg	Pro	Ser	Thr	Ala	Pro	Arg	Ser	Val	Thr	Thr	Lys	Lys
				635					640					645
Pro	Pro	Leu	Ala	Gln	Lys	Pro	Ser	Val	Ala	Pro	Arg	Gln	Asp	His
				650					655					660
Gln	Leu	Asp	Lys	Lys	Asp	Gly	Ile	Tyr	Pro	Ala	Gln	Leu	Val	Asn
				665					670					675

Tyr

## (2) INFORMATION FOR SEQ ID NO:8:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2338

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: mouse

(ix) FEATURE: cDNA for Huntingtin-interacting protein - mHIP1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GGCACGAGGG	CTCATTCAGA	TCCCCAGCT	GCCCGAGAAT	CCACCCAACTT	50
CCTACGAGCC	TCGGCCCTGT	CAGAGCACAT	CAGTCCTGTG	GTGGTGATCCC	100
GGCAGAGGTG	TCATCCCCAG	ACAGTGAGCC	TGTCCTGGAG	AAGGATGACCT	150
CATGGACATG	GACGCCTCCC	AGCAGACTTT	GTTTGACAAC	AAGTTTGATGA	200

CGTCTTTGGC	AGCTCATTGA	GCAGCGACCC	TTTCAATTTC	AACAATCAAAA	250
TGGCGTGAAC	AAGGACGAGA	AGGACCACTT	GATTGAACGC	CTGTACAGAGA	300
GATCAGTGGA	CTGACAGGGC	AGCTGGACAA	CATGAAGATT	GAGAGCCAGCG	350
GGCCATGCTG	CAGCTGAAGG	GTCGAGTGAG	TGAGCTGGAG	GCAGAGCTAGC	400
AGAGCAGCAG	CACTTGGGCC	GGCAGGCTAT	GGATGACTGC	GAGTTCCTGCG	450
CACTGAGCTG	GATGAACTGA	AGAGGCAGCG	AGAGGACACG	GAGAAGGCACA	500
GCGCAGCCTG	ACTGAGATAG	AAAGAAAGGC	CCAGGCTAAT	GAACAGAGGTA	550
TAGCAAGTTA	AAAGAGAAGT	ACAGTGAAC	GGTGCAGAAC	CATGCTGACCT	600
GCTGCGGAAG	AACGCAGAGG	TGACCAAACA	GGTGTCCGTG	GCCCGGCAAGC	650
CCAGGTGGAT	TTGGAAAGAG	AGAAAAAAGA	GCTAGCAGAT	TCCTTTGCAC	700
GTGTAAGTGA	CCAGGCCCCAG	CGGAAGACTC	AAGAGCAACA	GGATGTTCTA	750
GAGAACCTGA	AGCATGAACT	GGCCACCAGC	AGACAGGAGC	TGCAGGTCCT	800
CCACAGCAAC	CTGGAAACCT	CTGCCCAGTC	AGAAGCGAAA	TGGCTGACAC	850
AGATCGCCGA	GTTGGAGAAG	GAACAAGGCA	GCTTGGCGAC	TGTTGCAGCT	900
CAGAGAGAGG	AAGAGTTATC	AGCCCTCCGA	GACCAGCTGG	AAAGCACCCA	950
GATCAAGCTG	GCTGGGGCCC	AGGAATCCAT	GTGCCAGCAG	GTGAAGGACC	1000
AGAGGAAAAC	CCTCTTGGCA	GGGATCAGGA	AGGCTGCGGA	GCGTGAGATA	1050
CAGGAGGCGC	TGAGCCAGCT	TGAGGAACCC	ACCCTCATCA	GCTGTGCAGG	1100
ATCCACAGAT	CACCTTCTCT	CCAAAGTCAG	CTCCGTTTCC	AGCTGCCTCG	1150
AGCAACTGGA	AAAGAACGGC	AGCCAGTATC	TGGCCTGCCC	AGAAGATATT	1200
AGTGAGCTTC	TGCACTCGAT	CACCCTGCTT	GCCCACCTGA	CCGGTGACAC	1250
TGTCATCCAG	GGGAGTGCCA	CCAGCCTCCG	GGCCCCACCG	GAGCCAGCCG	1300
ACTCGTTGAC	GGAGGCCTGT	AGGCAGTATG	GCAGAGAAAC	CCTGGCCTAT	1350
CTGTCCTCCC	TGGAGGAAGA	GGGAAGTGTG	GAGAATGCTG	ACGTCACAGC	1400
CCTTAGGAAT	TGCCTCAGCA	GGGTCAAGAC	CCTTGGCGAG	GAGCTGCTGC	1450
CCAGGGGCCCT	GGACATCAAG	CAGGAAGAGC	TGGGTGACCT	GGTGGACAAG	1500
GAGATGGCAG	CCACTTCAGC	TGCCATTGAA	GCTGCCACCA	CCCGGATAGA	1550
GGAAATTCTC	AGTAAGTCCC	GAGCAGGAGA	CACGGGAGTC	AAGCTGGAGG	1600
TGAATGAGAG	GATCCTGGGT	TCCTGTACCA	GCCTGATGCA	GGCCATCAAG	1650
GTGCTCGTTG	TGGCCTCCAA	GGACCTCCAG	AAGGAGATAG	TGGAGAGTGG	1700
CAGGGGTAGT	GCATCCCCTA	AAGAATTTTA	CGCCAAGAAC	TCTCGGTGGA	1750
CGGAAGGGCT	GATATCCGCC	TCCAAAGCTG	TTGGTTGGGG	AGCTACCATC	1800
ATGGTGGATG	CTGCTGATCT	TGTGGTCCAA	GGCAAAGGGA	AGTTCGAGGA	1850
GCTGATGGTG	TGTTACGCG	AGATTGCTGC	CAGTACTGCC	CAGCTCGTGG	1900
CTGCATCCAA	GGTGAAAGCG	AACAAGGGCA	GCCTCAATCT	GACCCAGCTG	2000
CAGCAGGCCCT	CTCGAGGAGT	GAACCAGGCC	ACAGCCGCTG	TGGTGGCCTC	2050
AACCATTCTT	GGCAAATCTC	AGATTGAGGA	AACAGACAGT	ATGGACTTCT	2100
CAAGCATGAC	ACTGACCCAG	ATCAAGCGCC	AGGAGATGGA	TTCCCAGGTT	2150
AGGGTGCTGG	AGCTGGAAAA	TGACCTGCAG	AAGGAGCGTC	AGAAACTAGG	2200
AGAGCTACGG	AAGAAACACT	ACGAGCTGGA	GGGCGTGGCT	GAGGGCTGGG	2250
AGGAAGGGAC	AGAAGCATCA	CCGTCTACTG	TCCAAGAAGC	AATACCGGAC	2300
AAAGAGTAGA	GCCAAGCCGA	CACCCACAC	ATCAGAAA		2338

## (2) INFORMATION FOR SEQ ID NO: 9:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 676

(B) TYPE: protein

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: no

(vi) ORIGINAL SOURCE:

(A) ORGANISM: mouse

(ix) FEATURE: Huntingtin-interacting protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

Ala	Arg	Gly	Leu	Ile	Gln	Ile	Pro	Gln	Leu	Pro	Glu	Asn	Pro	Pro	5	10	15
Asn	Phe	Leu	Arg	Ala	Ser	Ala	Leu	Ser	Glu	His	Ile	Ser	Pro	Val	20	25	30
Val	Val	Ile	Pro	Ala	Glu	Val	Ser	Ser	Pro	Asp	Ser	Glu	Pro	Val	35	40	45
Leu	Glu	Lys	Asp	Asp	Leu	Met	Asp	Met	Asp	Ala	Ser	Gln	Gln	Thr	50	55	60
Leu	Phe	Asp	Asn	Lys	Phe	Asp	Asp	Val	Phe	Gly	Ser	Ser	Leu	Ser	65	70	75
Ser	Asp	Pro	Phe	Asn	Phe	Asn	Asn	Gln	Asn	Gly	Val	Asn	Lys	Asp	80	85	90
Glu	Lys	Asp	His	Leu	Ile	Glu	Arg	Leu	Tyr	Arg	Glu	Ile	Ser	Gly	95	100	105
Leu	Thr	Gly	Gln	Leu	Asp	Asn	Met	Lys	Ile	Glu	Ser	Gln	Arg	Ala	110	115	120
Met	Leu	Gln	Leu	Lys	Gly	Arg	Val	Ser	Glu	Leu	Glu	Ala	Glu	Leu	125	130	135
Ala	Glu	Gln	Gln	His	Leu	Gly	Arg	Gln	Ala	Met	Asp	Asp	Cys	Glu	140	145	150
Phe	Leu	Arg	Thr	Glu	Leu	Asp	Glu	Leu	Lys	Arg	Gln	Arg	Glu	Asp	155	160	165
Thr	Glu	Lys	Ala	Gln	Arg	Ser	Leu	Thr	Glu	Ile	Glu	Arg	Lys	Ala	170	175	180
Gln	Ala	Asn	Glu	Gln	Arg	Tyr	Ser	Lys	Leu	Lys	Glu	Lys	Tyr	Ser	185	190	195
Glu	Leu	Val	Gln	Asn	His	Ala	Asp	Leu	Leu	Arg	Lys	Asn	Ala	Glu	200	205	210
Val	Thr	Lys	Gln	Val	Ser	Val	Ala	Arg	Gln	Ala	Gln	Val	Asp	Leu	215	220	225
Glu	Arg	Glu	Lys	Lys	Glu	Leu	Ala	Asp	Ser	Phe	Ala	Arg	Val	Ser			

	230	235	240
Asp Gln Ala Gln Arg Lys Thr Gln Glu Gln Gln Asp Val Leu Glu	245	250	255
Asn Leu Lys His Glu Leu Ala Thr Ser Arg Gln Glu Leu Gln Val	260	265	270
Leu His Ser Asn Leu Glu Thr Ser Ala Gln Ser Glu Ala Lys Trp	275	288	285
Leu Thr Gln Ile Ala Glu Leu Glu Lys Glu Gln Gly Ser Leu Ala	290	295	300
Thr Val Ala Ala Gln Arg Glu Glu Glu Leu Ser Ala Leu Arg Asp	305	310	315
Gln Leu Glu Ser Thr Gln Ile Lys Leu Ala Gly Ala Gln Glu Ser	320	325	330
Met Cys Gln Gln Val Lys Asp Gln Arg Lys Thr Leu Leu Ala Gly	335	340	345
Ile Arg Lys Ala Ala Glu Arg Glu Ile Gln Glu Ala Leu Ser Gln	350	355	360
Leu Glu Glu Pro Thr Leu Ile Ser Cys Ala Gly Ser Thr Asp His	365	370	375
Leu Leu Ser Lys Val Ser Ser Val Ser Ser Cys Leu Glu Gln Leu	380	385	390
Glu Lys Asn Gly Ser Gln Tyr Leu Ala Cys Pro Glu Asp Ile Ser	395	400	405
Glu Leu Leu His Ser Ile Thr Leu Leu Ala His Leu Thr Gly Asp	410	415	420
Thr Val Ile Gln Gly Ser Ala Thr Ser Leu Arg Ala Pro Pro Glu	425	430	435
Pro Ala Asp Ser Leu Thr Glu Ala Cys Arg Gln Tyr Gly Arg Glu	440	445	450
Thr Leu Ala Tyr Leu Ser Ser Leu Glu Glu Glu Gly Thr Val Glu	455	460	465
Asn Ala Asp Val Thr Ala Leu Arg Asn Cys Leu Ser Arg Val Lys	470	475	480

Thr	Leu	Gly	Glu	Glu	Leu	Leu	Pro	Arg	Gly	Leu	Asp	Ile	Lys	Gln
				485					490					495
Glu	Glu	Leu	Gly	Asp	Leu	Val	Asp	Lys	Glu	Met	Ala	Ala	Thr	Ser
				500					505					510
Ala	Ala	Ile	Glu	Ala	Ala	Thr	Thr	Arg	Ile	Glu	Glu	Ile	Leu	Ser
				515					520					525
Lys	Ser	Arg	Ala	Gly	Asp	Thr	Gly	Val	Lys	Leu	Glu	Val	Asn	Glu
				530					535					540
Arg	Ile	Leu	Gly	Ser	Cys	Thr	Ser	Leu	Met	Gln	Ala	Ile	Lys	Val
				545					550					555
Leu	Val	Val	Ala	Ser	Lys	Asp	Leu	Gln	Lys	Glu	Ile	Val	Glu	Ser
				560					565					570
Gly	Arg	Gly	Ser	Ala	Ser	Pro	Lys	Glu	Phe	Tyr	Ala	Lys	Asn	Ser
				575					588					585
Arg	Trp	Thr	Glu	Gly	Leu	Ile	Ser	Ala	Ser	Lys	Ala	Val	Gly	Trp
				590					595					600
Gly	Ala	Thr	Ile	Met	Val	Asp	Ala	Ala	Asp	Leu	Val	Val	Gln	Gly
				605					610					615
Lys	Gly	Lys	Phe	Glu	Glu	Leu	Met	Val	Cys	Ser	Arg	Glu	Ile	Ala
				620					625					630
Ala	Ser	Thr	Ala	Gln	Leu	Val	Ala	Ala	Ser	Lys	Val	Lys	Ala	Asn
				635					640					645
Lys	Gly	Ser	Leu	Asn	Leu	Thr	Gln	Leu	Gln	Gln	Ala	Ser	Arg	Gly
				650					655					660
Val	Asn	Gln	Ala	Thr	Ala	Ala	Val	Val	Ala	Ser	Thr	Ile	Ser	Gly
				665					670					675
Lys	Ser	Gln	Ile	Glu	Glu	Thr	Asp	Ser	Met	Asp	Phe	Ser	Ser	Met
				680					685					690
Thr	Leu	Thr	Gln	Ile	Lys	Arg	Gln	Glu	Met	Asp	Ser	Gln	Val	Arg
				695					700					705
Val	Leu	Glu	Leu	Glu	Asn	Asp	Leu	Gln	Lys	Glu	Arg	Gln	Lys	Leu
				710					715					720
Gly	Glu	Leu	Arg	Lys	Lys	His	Tyr	Glu	Leu	Glu	Gly	Val	Ala	Glu
				725					730					735

Gly Trp Glu Glu Gly Thr Glu Ala Ser Pro Ser Thr Val Gln Glu  
740 745 750

Ala Ile Pro Asp Lys Glu  
755

## (2) INFORMATION FOR SEQ ID NO:10:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3964

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(vi) ORIGINAL SOURCE:

(A) ORGANISM: mouse

(ix) FEATURE: cDNA for Huntingtin-interacting protein - mHIP1a

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

GGCACGAGGC	GGCGCGCGGC	CTCCGTGTGC	CTAGGCTTGA	GGCGGGCGGT	50
GACGCCTCAT	TCGCGCGGAG	CCGGGCCGGG	ACACGGTCGG	CGGCAGCATG	100
AACAGCATCA	AGAATGTGCC	GGCGCGGGTG	CTGAGCCGCA	GGCCGGGCCA	150
CAGCCTAGAG	GCCGAGCGCG	AGCAGTTCGA	CAAGACGCAG	GCCATCAGTA	200
TCAGCAAAGC	CATCAACAGC	CAGGAGGCC	CAGTGAAGGA	GAAGCATGCC	250
CGGCGTATCA	TCCTGGGCAC	GCATCATGAG	AAGGGAGCCT	TCACCTTCTG	300
GTCCTATGCC	ATCGGCCTGC	CGCTGTCCAG	CAGCTCCATC	CTCAGCTGGA	350
AGTTCTGTCA	CGTCCTTCAC	AAGGTCCTCC	GGGACGGACA	CCCCAACGTC	400
CTGCATGACT	ATCAGCGGTA	CCGGAGCAAC	ATACGTGAGA	TCGGTGACTT	450
GTGGGGCCAC	CTTCGTGACC	AGTATGGACA	CCTGGTGAAT	ATCTATACCA	500
AACTGTTGCT	GACTAAGATC	TCCTTCCACC	TTAAGCACCC	CCAGTTTCTT	550
GCAGGCCTGG	AGGTAACAGA	TGAGGTGTTG	GAGAAGGCGG	CGGGAAGTGA	600
TGTCAACAAC	ATTTTTCAGC	TTACCGTGGA	GATGTTTGAC	TACATGGACT	650
GTGAACTGAA	GCTTTCTGAG	TCAGTTTTC	GGCAGCTCAA	CACGGCCATC	700
GCAGTGTCCT	AGATGTCTTC	TGGCCAGTGT	CGCCTAGCGC	CGCTCATCCA	750
GGTCATTTCAG	GACTGCAGCC	ACCTGTACCA	CTACACAGTG	AAGCTCATGT	800
TTAAGCTGCA	CTCCTGTCTC	CCGGCAGACA	CCCTGCAAGG	CCACAGGGAT	850
CGGTTCCACG	AGCAGTTCCA	CAGCCTCAAA	AACTTCTTCC	GCCGGGCTTC	900
AGACATGCTG	TACTTCAAGA	GGCTCATCCA	GATCCCGCGG	CTGCCTGAGG	950
GACCCCCCAA	TTTCCTGCGG	GCTTCAGCCC	TGGCTGAGCA	CATCAAGCCG	1000
GTGGTGGTGA	TTCCCGAGGA	GGCCCCAGAG	GAAGAGGAGC	CTGAGAACCT	1050
AATTGAAATC	AGCAGTGCGC	CCCCTGCTGG	GGAGCCAGTG	GTGGTGGCTG	1100
ACCTCTTTGA	TCAGACCTTT	GGACCCCCCA	ATGGCTCCAT	GAAGGATGAC	1150
AGGGACCTCC	AAATCGAGAA	CTTGAAGAGA	GAGGTGGAGA	CCCTCCGTGC	1200
TGAGCTGGAG	AAGATTAAGA	TGGAGGCACA	GCGGTACATC	TCCCAGCTGA	1250
AGGGCCAGGT	GAATGGCCTG	GAGGCAGAGC	TGGAGGAGCA	GCGCAAGCAG	1300
AAGCAGAAGG	CCCTGGTGGA	CAACGAGCAG	CTGCGCCACG	AGCTGGCCCCA	1350
GCTCAAGGCC	CTGCAGCTGG	AGGGCGCCCG	CAACCAGGGC	CTTCGAGAGG	1400
AAGCAGAGAG	GAAGGCCAGT	GCCACGGAGG	CACGCTACAG	CAAGCTGAAG	1450
GAGAAACACA	GCGAACTCAT	TAACACGCAC	GCCGAGCTGC	TCAGGAAGAA	1500



CGCAGACACG	GCCAAGCAGC	TGACAGTGAC	ACAGCAGAGC	CAGGAGGAGG	1550
TGGCACGGGT	AAAGGAACAG	CTGGCCTTCC	AGATGGAGCA	AGCGAAGCGT	1600
GAGTCTGAGA	TGAAGATGGA	AGAGCAGAGC	GACCAGTTGG	AGAAGCTCAA	1650
GAGGGAGCTG	GCGGCCAGGG	CAGGAGAGCT	GGCCCGTGCG	CAGGAGGCCC	1700
TGAGCCGCAC	AGAACAGAGT	GGGTCAGAGC	TGAGCTCACG	GCTGGACACA	1750
CTGAACGCGG	AGAAGGAAGC	CCTGAGTGGA	GTCGTTTCGGC	AGCGTGAGGC	1800
AGAGCTGCTG	GCCGCTCAGA	GCCTGGTGCG	GGAGAAGGAG	GAGGCGCTTA	1850
GCCAAGAGCA	GCAGCGGAGC	TCCCAGGAGA	AGGGCGAGCT	ACGGGGGCGAG	1900
CTGGCAGAAA	AGGAGTCTCA	GGAGCAGGGG	CTTCGGCAGA	AGCTGCTGGA	1950
TGAGCAGTTG	GCGGTGTTGC	GAAGTGCAGC	CGCCGAGGCA	GAGGCCATCC	2000
TACAGGATGC	AGTGAGCAAG	CTGGACGACC	CCCTGCACCT	CCGCTGCACC	2050
AGCTCCCCAG	ACTACTTGGT	GAGCCGGGCT	CAGGCAGCCC	TGGACAGCGT	2100
GAGCGGCCTG	GAGCAGGGCC	ACACCCAGTA	CCTGGCTTCC	TCCGAAGATG	2150
CTTCTGCCCT	GGTGGCAGCG	CTGACCCGCT	TCTCCCATT	GGCTGCGGAC	2200
ACCATTGTCA	ATGGTGCCGC	CACCTCCAC	CTGGCCCCCA	CCGACCCCGC	2250
CGACCGCCTG	ATGGACACAT	GCAGGGAGTG	TGGAGCCCCG	GCTCTGGAGC	2300
TGGTGGGACA	GCTGCAAGAC	CAGACAGTGC	TACGGAGGGC	TCAGCCCAGC	2350
CTGATGCGGG	CCCCCTGCA	GGGCATTCTG	CAGTTGGGCC	AGGACTTGAA	2400
GCCTAAGAGC	CTGGATGTAC	GGCAAGAGGA	GCTAGGGGCC	ATGGTGGACA	2450
AGGAGATGGC	GGCCACCTCG	GCAGCCATTG	AGGACGCTGT	GCGGAGGATC	2500
GAGGACATGA	TGAGCCAGGC	CCGCCACGAG	AGCTCAGGCG	TGAAACTGGA	2550
GGTGAATGAG	AGGATCCTCA	ACTCCTGCAC	AGACCTGATG	AAGGCTATCC	2600
GGCTCCTGGT	GATGACCTCC	ACCAGCCTGC	AGAAGGAAAT	TGTGGAGAGC	2650
GGCAGGGGGG	CAGCAACGCA	GCAGGAATTT	TATGCCAAGA	ATTCACGGTG	2700
GACTIONAGGC	CTCATCTCAG	CCTCTAAGGC	AGTGGGCTGG	GGAGCCACAC	2750
AGCTGGTGGA	GTCAGCTGAC	AAGGTTGTGC	TTCACATGGG	CAAATACGAG	2800
GAACCTCATCG	TCTGCTCCCA	TGAGATTGCG	GCCAGCACGG	CCCAGCTGGT	2850
GGCAGCCTCG	AAGGTGAAAG	CCAACAAGAA	CAGTCCCCAC	TTGAGCCGCC	2900
TGCAGGAATG	TTCCCGCACT	GTCAACGAGA	GGGCTGCCAA	CGTCGTGGCC	2950
TCCACCAAAT	CTGGCCAGGA	GCAGATTGAG	GACAGAGACA	CCATGGATTT	3000
CTCTGGCCTG	TCCCTCATCA	AGTTGAAGAA	GCAGGAGATG	GAGACACAGG	3050
TGCGAGTCTT	GGAGCTGGAG	AAGACACTAG	AGGCAGAGCG	TGTCCGGCTC	3100
GGGGAGCTTC	GGAAACAGCA	CTATGTACTG	GCTGGGGGGA	TGGGAACACC	3150
TAGCGAAGAA	GAACCCAGCA	GACCCAGCCC	AGCTCCCCGA	AGTGGGGCCA	3200
CTAAGAAGCC	ACCGCTGGCC	CAGAAACCCA	GCATAGCCCC	CAGGACAGAC	3250
AACCAGCTCGA	CAAAAAGGAT	GGTGTCTACC	CAGCTCAACT	TGTGAACTAC	3300
TAGGCCCTTAA	GGTGTTTCAGC	AGGATGGCTG	GTGGTTGTGC	CTGGGCTTCA	3350
TGTGGCTGTCT	GGCAGTGGTC	AAGGGGCCTC	TGAGAAGCCT	CCAACCTCTG	3400
CCCAAGGGGCC	TAGTCTGTGG	GACAGTTCAT	CTGGATGTGA	ATCTATTTAT	3450
CTTAAGTAGGA	ACTGCCTCGA	GCAGCTGGGA	CCCAGCAGGC	CTGAGCCACA	3500
AATCTGCAGCG	GACATCAGAG	ATAGTCTGAA	TGCTGCGAGG	TATTTCTTTC	3550
TTCGTAAGTTT	AGTCAGCACA	CTGGGAAAAG	GTCACATAAG	CCAGGAGCCT	3600
CCTTGTTCTCTG	GACTCAAAAAG	TCTGAGGCCT	TAAGTGAACA	ACAGAAAGAG	3650
GGTCCCTGCTG	GCTACCAGGG	ATAAGGGGAT	GACCTGTGAC	CCTTGAGCCA	3700
GGGAGAGCAGG	TAAGCTGGGT	GGTGTCATCA	CCTGGGGGCC	TGGTGCTAGG	3750
GCATCCATGCT	GGGAGCCCCA	GGAGACCAGG	CTTTGTGTGG	GAGCCTGGCA	3800
TCATCGTGGCT	GGGGCAGCCC	CTGCTCAGGT	GCTGTCTCTG	CCCGTGACCT	3850
TGAAGCCACCC	TCCCCCGTA	CAGTTTTTCCA	TTCTCCTGGC	TACTAGTGTG	3900
GCTGTTTCATTG	CCTACCTTGA	TGAGTAGATT	TCAGCCCTCC	TAAAGCTGGG	3950
GCCTTTCTCTCG	TGCC				3964

## (2) INFORMATION FOR SEQ ID NO: 11:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 676

(B) TYPE: protein

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: no

(vi) ORIGINAL SOURCE:

(A) ORGANISM: mouse

(ix) FEATURE: Huntingtin-interacting protein -mHIP1a

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Met	Asn	Ser	Ile	Lys	Asn	Val	Pro	Ala	Arg	Val	Leu	Ser	Arg	Arg	5	10	15
Pro	Gly	His	Ser	Leu	Glu	Ala	Glu	Arg	Glu	Gln	Phe	Asp	Lys	Thr	20	25	30
Gln	Ala	Ile	Ser	Ile	Ser	Lys	Ala	Ile	Asn	Ser	Gln	Glu	Ala	Pro	35	40	45
Val	Lys	Glu	Lys	His	Ala	Arg	Arg	Ile	Ile	Leu	Gly	Thr	His	His	50	55	60
Glu	Lys	Gly	Ala	Phe	Thr	Phe	Trp	Ser	Tyr	Ala	Ile	Gly	Leu	Pro	65	70	75
Leu	Ser	Ser	Ser	Ser	Ile	Leu	Ser	Trp	Lys	Phe	Cys	His	Val	Leu	80	85	90
His	Lys	Val	Leu	Arg	Asp	Gly	His	Pro	Asn	Val	Leu	His	Asp	Tyr	95	100	105
Gln	Arg	Tyr	Arg	Ser	Asn	Ile	Arg	Glu	Ile	Gly	Asp	Leu	Trp	Gly	110	115	120
His	Leu	Arg	Asp	Gln	Tyr	Gly	His	Leu	Val	Asn	Ile	Tyr	Thr	Lys	125	130	135
Leu	Leu	Leu	Thr	Lys	Ile	Ser	Phe	His	Leu	Lys	His	Pro	Gln	Phe	140	145	150
Pro	Ala	Gly	Leu	Glu	Val	Thr	Asp	Glu	Val	Leu	Glu	Lys	Ala	Ala	155	160	165
Gly	Thr	Asp	Val	Asn	Asn	Ile	Phe	Gln	Leu	Thr	Val	Glu	Met	Phe	170	175	180
Asp	Tyr	Met	Asp	Cys	Glu	Leu	Lys	Leu	Ser	Glu	Ser	Val	Phe	Arg			

				185					190					195
Gln	Leu	Asn	Thr	Ala	Ile	Ala	Val	Ser	Gln	Met	Ser	Ser	Gly	Gln
				200						205				210
Cys	Arg	Leu	Ala	Pro	Leu	Ile	Gln	Val	Ile	Gln	Asp	Cys	Ser	His
				215						220				225
Leu	Tyr	His	Tyr	Thr	Val	Lys	Leu	Met	Phe	Lys	Leu	His	Ser	Cys
				230						235				240
Leu	Pro	Ala	Asp	Thr	Leu	Gln	Gly	His	Arg	Asp	Arg	Phe	His	Glu
				245						250				255
Gln	Phe	His	Ser	Leu	Lys	Asn	Phe	Phe	Arg	Arg	Ala	Ser	Asp	Met
				260						265				270
Leu	Tyr	Phe	Lys	Arg	Leu	Ile	Gln	Ile	Pro	Arg	Leu	Pro	Glu	Gly
				275						288				285
Pro	Pro	Asn	Phe	Leu	Arg	Ala	Ser	Ala	Leu	Ala	Glu	His	Ile	Lys
				290						295				300
Pro	Val	Val	Val	Ile	Pro	Glu	Glu	Ala	Pro	Glu	Glu	Glu	Glu	Pro
				305						310				315
Glu	Asn	Leu	Ile	Glu	Ile	Ser	Ser	Ala	Pro	Pro	Ala	Gly	Glu	Pro
				320						325				330
Val	Val	Val	Ala	Asp	Leu	Phe	Asp	Gln	Thr	Phe	Gly	Pro	Pro	Asn
				335						340				345
Gly	Ser	Met	Lys	Asp	Asp	Arg	Asp	Leu	Gln	Ile	Glu	Asn	Leu	Lys
				350						355				360
Arg	Glu	Val	Glu	Thr	Leu	Arg	Ala	Glu	Leu	Glu	Lys	Ile	Lys	Met
				365						370				375
Glu	Ala	Gln	Arg	Tyr	Ile	Ser	Gln	Leu	Lys	Gly	Gln	Val	Asn	Gly
				380						385				390
Leu	Glu	Ala	Glu	Leu	Glu	Glu	Gln	Arg	Lys	Gln	Lys	Gln	Lys	Ala
				395						400				405
Leu	Val	Asp	Asn	Glu	Gln	Leu	Arg	His	Glu	Leu	Ala	Gln	Leu	Lys
				410						415				420
Ala	Leu	Gln	Leu	Glu	Gly	Ala	Arg	Asn	Gln	Gly	Leu	Arg	Glu	Glu
				425						430				435
Ala	Glu	Arg	Lys	Ala	Ser	Ala	Thr	Glu	Ala	Arg	Tyr	Ser	Lys	Leu

440										445			450		
Lys	Glu	Lys	His	Ser	Glu	Leu	Ile	Asn	Thr	His	Ala	Glu	Leu	Leu	
				455					460					465	
Arg	Lys	Asn	Ala	Asp	Thr	Ala	Lys	Gln	Leu	Thr	Val	Thr	Gln	Gln	
				470					475					480	
Ser	Gln	Glu	Glu	Val	Ala	Arg	Val	Lys	Glu	Gln	Leu	Ala	Phe	Gln	
				485					490					495	
Met	Glu	Gln	Ala	Lys	Arg	Glu	Ser	Glu	Met	Lys	Met	Glu	Glu	Gln	
				500					505					510	
Ser	Asp	Gln	Leu	Glu	Lys	Leu	Lys	Arg	Glu	Leu	Ala	Ala	Arg	Ala	
				515					520					525	
Gly	Glu	Leu	Ala	Arg	Ala	Gln	Glu	Ala	Leu	Ser	Arg	Thr	Glu	Gln	
				530					535					540	
Ser	Gly	Ser	Glu	Leu	Ser	Ser	Arg	Leu	Asp	Thr	Leu	Asn	Ala	Glu	
				545					550					555	
Lys	Glu	Ala	Leu	Ser	Gly	Val	Val	Arg	Gln	Arg	Glu	Ala	Glu	Leu	
				560					565					570	
Leu	Ala	Ala	Gln	Ser	Leu	Val	Arg	Glu	Lys	Glu	Glu	Ala	Leu	Ser	
				575					588					585	
Gln	Glu	Gln	Gln	Arg	Ser	Ser	Gln	Glu	Lys	Gly	Glu	Leu	Arg	Gly	
				590					595					600	
Gln	Leu	Ala	Glu	Lys	Glu	Ser	Gln	Glu	Gln	Gly	Leu	Arg	Gln	Lys	
				605					610					615	
Leu	Leu	Asp	Glu	Gln	Leu	Ala	Val	Leu	Arg	Ser	Ala	Ala	Ala	Glu	
				620					625					630	
Ala	Glu	Ala	Ile	Leu	Gln	Asp	Ala	Val	Ser	Lys	Leu	Asp	Asp	Pro	
				635					640					645	
Leu	His	Leu	Arg	Cys	Thr	Ser	Ser	Pro	Asp	Tyr	Leu	Val	Ser	Arg	
				650					655					660	
Ala	Gln	Ala	Ala	Leu	Asp	Ser	Val	Ser	Gly	Leu	Glu	Gln	Gly	His	
				665					670					675	
Thr	Gln	Tyr	Leu	Ala	Ser	Ser	Glu	Asp	Ala	Ser	Ala	Leu	Val	Ala	
				680					685					690	
Ala	Leu	Thr	Arg	Phe	Ser	His	Leu	Ala	Ala	Asp	Thr	Ile	Val	Asn	

695										700			705		
Gly	Ala	Ala	Thr	Ser	His	Leu	Ala	Pro	Thr	Asp	Pro	Ala	Asp	Arg	
710										715			720		
Leu	Met	Asp	Thr	Cys	Arg	Glu	Cys	Gly	Ala	Arg	Ala	Leu	Glu	Leu	
725										730			735		
Val	Gly	Gln	Leu	Gln	Asp	Gln	Thr	Val	Leu	Arg	Arg	Ala	Gln	Pro	
740										745			750		
Ser	Leu	Met	Arg	Ala	Pro	Leu	Gln	Gly	Ile	Leu	Gln	Leu	Gly	Gln	
755										760			765		
Asp	Leu	Lys	Pro	Lys	Ser	Leu	Asp	Val	Arg	Gln	Glu	Glu	Leu	Gly	
770										775			780		
Ala	Met	Val	Asp	Lys	Glu	Met	Ala	Ala	Thr	Ser	Ala	Ala	Ile	Glu	
785										790			795		
Asp	Ala	Val	Arg	Arg	Ile	Glu	Asp	Met	Met	Ser	Gln	Ala	Arg	His	
800										805			810		
Glu	Ser	Ser	Gly	Val	Lys	Leu	Glu	Val	Asn	Glu	Arg	Ile	Leu	Asn	
815										820			825		
Ser	Cys	Thr	Asp	Leu	Met	Lys	Ala	Ile	Arg	Leu	Leu	Val	Met	Thr	
830										835			840		
Ser	Thr	Ser	Leu	Gln	Lys	Glu	Ile	Val	Glu	Ser	Gly	Arg	Gly	Ala	
845										850			855		
Ala	Thr	Gln	Gln	Glu	Phe	Tyr	Ala	Lys	Asn	Ser	Arg	Trp	Thr	Glu	
860										865			870		
Gly	Leu	Ile	Ser	Ala	Ser	Lys	Ala	Val	Gly	Trp	Gly	Ala	Thr	Gln	
875										888			885		
Leu	Val	Glu	Ser	Ala	Asp	Lys	Val	Val	Leu	His	Met	Gly	Lys	Tyr	
890										895			900		
Glu	Glu	Leu	Ile	Val	Cys	Ser	His	Glu	Ile	Ala	Ala	Ser	Thr	Ala	
905										910			915		
Gln	Leu	Val	Ala	Ala	Ser	Lys	Val	Lys	Ala	Asn	Lys	Asn	Ser	Pro	
920										925			930		
His	Leu	Ser	Arg	Leu	Gln	Glu	Cys	Ser	Arg	Thr	Val	Asn	Glu	Arg	
935										940			945		
Ala	Ala	Asn	Val	Val	Ala	Ser	Thr	Lys	Ser	Gly	Gln	Glu	Gln	Ile	

950										955					960				
Glu	Asp	Arg	Asp	Thr	Met	Asp	Phe	Ser	Gly	Leu	Ser	Leu	Ile	Lys					
965										970					975				
Leu	Lys	Lys	Gln	Glu	Met	Glu	Thr	Gln	Val	Arg	Val	Leu	Glu	Leu					
980										985					990				
Glu	Lys	Thr	Leu	Glu	Ala	Glu	Arg	Val	Arg	Leu	Gly	Glu	Leu	Arg					
995										1100					1105				
Lys	Gln	His	Tyr	Val	Leu	Ala	Gly	Gly	Met	Gly	Thr	Pro	Ser	Glu					
1110										1115					1120				
Glu	Glu	Pro	Ser	Arg	Pro	Ser	Pro	Ala	Pro	Arg	Ser	Gly	Ala	Thr					
1125										1130					1135				
Lys	Lys	Pro	Pro	Leu	Ala	Gln	Lys	Pro	Ser	Ile	Ala	Pro	Arg	Thr					
1140										1145					1150				
Asp	Asn	Gln	Leu	Asp	Lys	Lys	Asp	Gly	Val	Tyr	Pro	Ala	Gln	Leu					
1155										1160					1165				
Val Asn Tyr																			

## (2) INFORMATION FOR SEQ ID NO:12:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other DNA

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

GAAGATACCC CACCAAAC 18

## (2) INFORMATION FOR SEQ ID NO:13:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 35

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other DNA

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

GCTTGACAGT GTAGTCATAA AGGTGGCTGC AGTCC 35

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other DNA

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

GGACATGTCC AGGGAGTTGA ATAC 24

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 41

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: yes

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

CUACUACUAC UACUAGGCCA CGCGTCGACT AGTACGGGII GGGIIGGGII G 41

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 516

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(x) FEATURE: exon 1 of HIP1

(xi)SEQUENCE DESCRIPTION: SEQ ID NO: 16:

TCTGTGGAAG	GTTTGGAGGG	GAGAGAGGGG	CAGCTGGATG	CTCTTGGGCC	ACGGTCGCCC	60
CTGATCTCTG	CGCCTCTTCC	TCCTGCTCCG	GGAGAAATAA	TGTTTCCCTG	GGGGATGAAA	120
GCATCTCTTT	GTGCGGGCTT	TAATTGCCAT	GTTGTTGTGC	CAAGGGAGTG	AGTGGCGGCG	180
GGACCAGCAG	CTGGGCACAG	CCAATGCCAG	GCAGTGGTGC	CCACTCCCTC	AGGACGCCCA	240
GCCAGCTGGC	TCCTGGGAGC	GCTGCCCACC	TCTGCCCCCA	GCTGGGCGCC	TGCAAGGAAC	300
CGACCACCCG	TGGGGCTGGG	GGAGGTTGGC	TGGAGGAGGA	GAAAGGGGCG	GGCTCTGGGA	360
GGGTCTCAGC	CACTCTCAGA	GGCTTATTCA	TCTCATCCTC	CTTTCCCTCC	CCCTTCTTGT	420
TTTTCAGACT	GTCAGCATCA	ATAAGGCCAT	TAATACGCAG	GAAGTGGCTG	TAAAGGAAAA	480
ACACGCCAGA	AATATCCTTT	GGATGTTGCT	TGGAAG			516

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 193

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii)MOLECULE TYPE: genomic DNA

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(x) FEATURE: exon 2 of HIP1

(xi)SEQUENCE DESCRIPTION: SEQ ID NO: 17:

TGTTTTCCAT	AACCCCCCCT	CACCGTGCAT	ACTGGGCACC	CACCATGAGA	AAGGGGCACA	60
GACCTTCTGG	TCTGTTGTCA	ACCGCCTGCC	TCTGTCTAGC	AACCCAGTGC	TCTGCTGGAA	120
GTTCTGCCAT	GTGTTCCACA	AACTCCTCCG	AGATGGACAC	CCGAACGTGA	GTTCTTGGGG	180
CTATGGGGTG	GCA					193

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 104

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii)MOLECULE TYPE: genomic DNA

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(x) FEATURE: exon 3 of HIP1

(xi)SEQUENCE DESCRIPTION: SEQ ID NO: 18:

GTGTTCTTTT	GCCCCTGCAG	GTCCTGAAGG	ACTCTCTGAG	ATACAGAAAT	GAATTGAGTG	60
ACATGAGCAG	GATGTGGGTG	AGTTTGGAGA	TGTA CT CAGG	AGCC		104

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:



- (A) LENGTH: 327  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear  
 (ii) MOLECULE TYPE: genomic DNA  
 (iii) HYPOTHETICAL: no  
 (iv) ANTI-SENSE: no  
 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: human  
 (x) FEATURE: exon 4 of HIP1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

```

AATTCCTGGC TGCAGATCTC TTGACTGTTA TGTTCTTGTT GTTGACTCTG TTTCCCCTCC   60
TCTTCCTAAA AGGGCCACCT GAGCGAGGGG TATGGCCAGC TGTGCAGCAT CTACCTGAAA   120
CTGCTAAGAA CCAAGATGGA GTACCACACC AAAGTGAGTC TCTGCGGACA GTTCTGCCGC   180
CACCGCCGCC TCCCCTGCTC CATCCCTTCA GCCCCTCCCT GGGCTCATTT GTCAGCTCTT   240
TCAGGTAATA GACAGCCAG GCTTCTGAGG AAGTGTGCAC ATCATGTACC CAAGCTGTGA   300
GAGAGGAAAG CCACCGCCAG GCCCAGC                                     327
  
```

## (2) INFORMATION FOR SEQ ID NO:21:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 331  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear  
 (ii) MOLECULE TYPE: genomic DNA  
 (iii) HYPOTHETICAL: no  
 (iv) ANTI-SENSE: no  
 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: human  
 (x) FEATURE: exon 5 of HIP1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

```

GGGCTCAAGC AATCCTCCCA CCTCGGCCTC CCAAGTAGCT GGGACCACAG GCGTGTGCCA   60
CCACGCCCCG CTGAGAGAGG GCTCTTCATG TCTTCTGCCC TGACTCCCTT CCTCTGCCTC   120
CCTTCCAGAA TCCCAGGTTT CCAGGCAACC TGCAGATGAG TGACCGCCAG CTGGACGAGG   180
CTGGAGAAAG TGACGTGAAC AACTTGTAAG TGGCTCCTGC CCTGAGCCCA GGGAGGGAGA   240
AAGCTTTTGT GAATGCTGAC ACTTCTCATA AGGGTCATGG AGGGCCTGAT GGGGGGAGGC   300
CGTGGCTGGG ATGGGGACCA AAGCCCCTGG G                                     331
  
```

## (2) INFORMATION FOR SEQ ID NO:22:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 470  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear  
 (ii) MOLECULE TYPE: genomic DNA  
 (iii) HYPOTHETICAL: no  
 (iv) ANTI-SENSE: no

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(x) FEATURE: exon 6 of HIP1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

ACTGTCGCTG	TCACGTGTTGA	CTTCACCAGG	CTGCATGGCC	ATAATACCCA	CAAGGCTAAG	60
ACTTGGAGCT	GGAGTTGTGT	GTGTGTTTGC	GCATGCACAT	GAGCATTTGA	GACTGGAGTA	120
GCGTAGAGCG	TGGGGGAGGG	GACAGGTAAC	AGACCGGCCT	CAGGCTGTGG	AGTGTAAAGCT	180
CTCTTTCTCT	TTGGGTCCAG	TTTCCAGTTA	ACAGTGGAGA	TGTTTGACTA	CCTGGAGTGT	240
GAACCAACC	TCTTCCAAAC	AGGTGAGTCT	CTTCCCTCCC	GTCTAACCCA	GGCTCTCATG	300
GGAACACCT	AATTCCTAGT	CCTCCTCTCC	CTGCAAAGTG	TGCAGCACAA	GGGGTAGGAA	360
AATGGAGACA	TTCACACCCC	ATCTCTGGTC	TCTCCAACCC	TCGTGCAGGG	AGGGACTGAA	420
CCTCTTCAGT	ATTTTTCTTT	TTAAGAGACA	AGGTCTCGGC	CGGGTGCAGT		470

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 565

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(x) FEATURE: exon 7 of HIP1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

TCTTCACCTG	TTTAATGGGG	ATACGTTTAC	CTATCTCATG	GGAGTGTTGT	GAAGGTTAAA	60
TGAATTAGAT	GAGGTAAAGC	ACGCACAGAA	TCGGTCCTTG	GTGTATGTTG	GACCCCTGCC	120
TCTGCCCCCT	TGAAGAGGCT	GCCTGTAATC	CCCTGGCTCT	ACCACCTTTC	TCCCTCACTT	180
TTATTTCTTA	GTATTCAACT	CCCTGGACAT	GTCCCGCTCT	GTGTCCGTGA	CGGCAGCAGG	240
GCAGTGCCGC	CTCGCCCCGC	TGATCCAGGT	CATCTTGGAC	TGCAGCCACC	TTTATGACTA	300
CACTGTCAAG	CTTCTCTTCA	AACTCCACTC	CTGTGAGTAC	CGCGGGCCAG	ATCTTCTTAC	360
ATGAGATTCA	GGCCAGAGGG	AGGATCCAG	CCTGAGGATG	TCCCCAGAGA	AACGCAGTCC	420
TTCTCAGTGC	CTTTGGCTGT	CTGCTTCTGT	TCCAAAAGGC	CCCGGAGCTT	CTGACCATTG	480
TGAGGATAAA	AGAGCAGGGC	CCAGGCTTTG	GTGACCCAG	TAAAGCCCCT	GGCTTGCCAC	540
TCTTGCGTCC	AGTGTTACAG	GATCT				565

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 233

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(x) FEATURE: exon 8 of HIP1

## (xi)SEQUENCE DESCRIPTION: SEQ ID NO: 24:

GGGACAGCTC	TAGGCCAGTC	GTGGCCCCCTG	GCAGTGCTGG	CCACATGCCC	CAGGGTAGCT	60
GGGCCCCCTCC	CCCTCGAGAG	CCCCGCTGTG	GCTTCCCTGC	CCTCTGGTCC	CCCTCCCCTC	120
TCACACTCTT	TCCAATTTCT	TCCAGGCCTC	CCAGCTGACA	CCCTGCAAGG	CCACCGGGAC	180
CGCTTCATGG	AGCAGTTTAC	AAAGTAAGTG	GTTCAAGTAA	CAGGAATGGA	GGT	233

## (2) INFORMATION FOR SEQ ID NO:25:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 578

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii)MOLECULE TYPE: genomic DNA

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(x) FEATURE: exons 9 and 10 of HIP1

## (xi)SEQUENCE DESCRIPTION: SEQ ID NO: 25:

TGAATCCCAG	CACCATGGAG	TTTATCTCCT	TGACAGCCTG	TGCCTTTGGG	CTGGGGAGGG	60
GGCAGGAAAG	CCAGGTGGCT	GCTCTGTCCC	CTACATGGGG	CTGATGAAGA	CACCCAGCAC	120
CCCTCAGGTC	CTTCTCCACC	CCTAGGTTGA	AAGATCTGTT	CTACCGCTCC	AGCAACCTGC	180
AGTACTTCAA	GCGGCTCATT	CAGATCCCCC	AGCTGCCTGA	GGTAAGCATG	CCCAACCACA	240
CACCCTCGGC	ACTGCAGAGG	CCCCAGGTAC	TCTCTTAAGG	GCCGGCGGGG	CCTGGCAAGC	300
AAGCACTATT	TGAGGATGTG	TCTCCGTCTT	CAGAACCCAC	CCAACTTCCT	GCGAGCCTCA	360
GCCCTGTCAG	AACATATCAG	CCCTGTGGTG	GTGATCCCTG	CAGAGGCCTC	ATCCCCCGAC	420
AGCGAGCCAG	TCCTAGAGAA	GGATGACCTC	ATGGACATGG	ATGCCTCTCA	GCAGGTGAGG	480
ACCACTTGGG	AGAGAAACTT	GGCCTTTCTT	CTCACCTGCA	AGTACAGGGG	AGAGGCTGGG	540
GGAGACCCTG	GCCAAAGCCC	ATTGACTCTA	ACCAGGTT			578

## (2) INFORMATION FOR SEQ ID NO:26:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 390

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii)MOLECULE TYPE: genomic DNA

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(x) FEATURE: exon 11 of HIP1

## (xi)SEQUENCE DESCRIPTION: SEQ ID NO: 26:

AAAAAAATTT	AAAAAATTAA	ACAGGTCTGA	ACCGTTTAAT	TCGAGAAAGG	GGGCATTCTC	60
CCATATCACT	CAACTGACCC	ACACACAGAA	TTCTCTGGCT	CTCTGACTTA	TTCTCACTCC	120
TTTTTGGTCA	ACCACAGAAT	TTATTTGACA	ACAAGTTTGA	TGACATCTTT	GGCAGTTCAT	180
TCAGCAGTGA	TCCCTTCAAT	TTCAACAGTC	AAAATGGTGT	GAACAAGGAT	GAGAAGTGAG	240
TCCAAGCTGG	GTTCAAGCAG	ATGGTTCAGG	AGCTAAGTTA	AGCCATGGTC	TGCCTCAAAA	300
CACTAACCAA	AGAGGAATTC	TTAATGATAC	TGGGGCTTCT	TAGATACAGA	ACATCTTGAA	360
GGGTGGGGG	CAATGGCTTA	TGCCTGTAAT				390

## (2) INFORMATION FOR SEQ ID NO:27:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 547  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: genomic DNA

## (iii) HYPOTHETICAL: no

## (iv) ANTI-SENSE: no

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: human

## (x) FEATURE: exon 12 of HIP1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

AAAATCAATA	ACCATGGATT	TATGAGTATT	AGATTAGTAT	CTGGTAACAT	TTAGAGTATA	60
ATTTATGGCA	TTTCAAAGAA	TTGTCCCCAA	ATTAATACCA	GCTTTTAATT	TCCTCCCCTG	120
AGCTCACAAT	TAAAAACAGA	GGGATAGAAG	CACTATGAAA	GCAAACATCAT	TCCCCTTCTC	180
TTCCCAGGGA	CCACTTAATT	GAGCGACTAT	ACAGAGAGAT	CAGTGGATTG	AAGGCACAGC	240
TAGAAAACAT	GAAGACTGAG	GTATAACTTG	GATCTGCTCT	GCCTTTGCGC	TTCACCAAAA	300
CACGGTAGAT	TTGAATGTTA	AATTTGCATC	ACACTAGCCA	GGCACAGTGG	CTCACACCTG	360
TAATCCTAGC	ACTTTGGGAG	GCCAAGGCAG	GAGGATTACC	TGAGGTCGGG	AGTTCGAGAC	420
CAGCCTGGGC	AACAGGGTGA	AACCCCGTC	TTCAATAAAA	ATGCAATAAT	TAGCCGGGTG	480
TGTTGGCAGG	CACCTGTAAT	CCCAGCTACT	CGGGAAGCTG	AGGCATGAGA	ATTGCTTGAA	540
CTTGGA						547

## (2) INFORMATION FOR SEQ ID NO:28:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 436  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: genomic DNA

## (iii) HYPOTHETICAL: no

## (iv) ANTI-SENSE: no

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: human

## (x) FEATURE: exon 13 of HIP1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

CCCCCAGCCA	CTCTAAAGAG	GACCACAATT	CCCCGGCCAT	CATCCCCTGT	TATTGTTGTT	60
GATTGAGGGG	CTCCTAATGA	CCAGATGGTC	CAACCCTCCT	GGGACGTGGA	GAGTTGACTT	120
AGGGGAATCA	GGTATTTACT	TGGAAGCATG	GTAGGACCCG	CTTCTCCGGC	CCATGCCCCG	180
GACCCGTGGC	AGTGGGCGGT	TGGCCTCATG	ACCGGAGTCC	CCCCACAGAG	CCAGCGGGTT	240
GTGCTGCAGC	TGAAGGGCCA	CGTCAGCGAG	CTGGAAGCAG	ATCTGGCCGA	GCAGCAGCAC	300
CTGCGGCAGC	AGGCGGCCGA	CGACTGTGAA	TTCTGCGGG	CAGAACTGGA	CGAGCTCAGG	360
AGGCAGCGGG	AGGACACCGA	GAAGGCTCAG	CGGAGCCTGT	CTGAGATAGA	AAGTGAGCGG	420
TGGGTGGGGG	CGGGG					436

## (2) INFORMATION FOR SEQ ID NO:29:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 469  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear  
(ii) MOLECULE TYPE: genomic DNA  
(iii) HYPOTHETICAL: no  
(iv) ANTI-SENSE: no  
(vi) ORIGINAL SOURCE:  
(A) ORGANISM: human  
(x) FEATURE: exon 14 of HIP1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

GACTTGAGCC	CAAGGAGGTC	AAGGCTGCAG	TGAACAGTGA	TTGTGCCACT	GCACCCAGC	60
CTGGGTGACA	GAGCAAGACT	GTCTCAAAAC	AAAACAAGGA	GGACCTTCTA	GGGACCCTGG	120
CTCATTGCAA	GGAAGGCAAG	GGTCCCTGCT	AGGTTAGACT	CCTCACCTTG	GTCCCTTTACA	180
ATACAGGGAA	AGCTCAAGCC	AATGAACAGC	GATATAGCAA	GCTAAAGGAG	AAGTACAGCG	240
AGCTGGTTCA	GAACCAACGCT	GACCTGCTGC	GGAAGGTAAG	ACCCTCAGCC	CCTGTCACCA	300
TCCTGCAGGC	CCTGCACCTC	TAGGGAGAGA	GCGGCTCAGG	CCTGTGGCTT	CCCCGGGGCC	360
AGCAACCCCT	ACATTGATCT	CTAAGGCATT	GCCGTCATCT	CGGGAACCAC	ACCTTTTCAG	420
GCTTCCTTGC	CTCTGTGTCT	TGGGCTGTGT	CCTGGGTGCC	AATCCCATG		469

## (2) INFORMATION FOR SEQ ID NO:30:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 359  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear  
(ii) MOLECULE TYPE: genomic DNA  
(iii) HYPOTHETICAL: no  
(iv) ANTI-SENSE: no  
(vi) ORIGINAL SOURCE:

- (A) ORGANISM: human  
(x) FEATURE: exon 15 of HIP1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

GGGTAGGAAA	GTGATTCCTG	TGTCTGACTC	TAGGGCACGC	ACAGCCTGAG	TATGATTGTC	60
CTAGAAGGAG	GATGTCCTCT	AAGCCTGGGA	TCTCCTGGTT	CAAGACACTG	TTCTTCTTTT	120
GCAGAATGCA	GAGGTGACCA	AACAGGTGTC	CATGGCCAGA	CAAGCCCAGG	TAGATTTGGA	180
ACGAGAGAAA	AAAGAGCTGG	AGGATTCGTT	GGAGCGCATC	AGTGACCAGG	GCCAGCGGAA	240
GGTGAGTGGG	ACGAGGAGCA	CTCGGAAAT	GAGGGAGGGG	GCTGTTGAGT	TGGTGGCGGG	300
GGCTTTGTGG	CCTTCTGCTC	CATGGGCAGT	TCTGTGGGTC	GGTTGGCATC	ACACAGCAG	359

## (2) INFORMATION FOR SEQ ID NO:31:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 209  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear  
(ii) MOLECULE TYPE: genomic DNA  
(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(x) FEATURE: exon 16 of HIP1

(xi)SEQUENCE DESCRIPTION: SEQ ID NO: 31:

GTTGATCGCT	TGGGACGTTT	TTACATTTTT	ATATTCTTTG	TCACTGTCAC	CCAGATCAGA	60
GTCCCTCTGT	TTTTCTTCTC	TTTCAGACTC	AAGAACAGCT	GGAAGTTCTA	GAGAGCTTGA	120
AGCAGGAAC	TGCCACAAGC	CAACGGGAGC	TTCAGGTTCT	GCAAGGCAGC	CTGGAAACTT	180
CTGCCAGGT	AAATACCTCC	TTTTTTTTT				209

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 485

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii)MOLECULE TYPE: genomic DNA

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(x) FEATURE: exon 17 of HIP1

(xi)SEQUENCE DESCRIPTION: SEQ ID NO: 32:

CCCCCACTGC	AATCAGTGTG	TCCCCGGGAG	GGAATCAGAG	TGGCAGGTTA	AAGAGCCATC	60
ACCTTCCCAG	TCCTTGCAAC	CCGGTGGTGG	GTTGGACCTC	TGGGAAGTAG	GGACTGTTTA	120
ACTCAACCAG	CGTCTCCCTC	TTTCCTTGTTG	GTCACCTTTG	CAGTCAGAAG	CAAACCTGGGC	180
AGCCGAGTTC	GCCGAGCTAG	AGAAGGAGCG	GGACAGCCTG	GTGAGTGGCG	CAGCTCATAG	240
GGAGGAGGAA	TTATCTGCTC	TTCGGAAAGA	ACTGCAGGAC	ACTCAGCTCA	AACTGGCCAG	300
CACAGAGGGT	CACGGACATG	GACACGAGCG	AGCACCTGTG	AATTCCCACC	GAGGGCCTCT	360
GCGCATGCAC	GGAGGCTGGG	AGGACCCCGG	GGCTGCTGAG	AAGGGGTTTG	GGGCCTTGGC	420
CTGATTGTGC	AGACATTCTG	TAGGTGTAAT	GCCAGCAGGC	CCTGCATTGC	CTGCAGAGTC	480
CATGA						485

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 468

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii)MOLECULE TYPE: genomic DNA

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(x) FEATURE: exon 18 of HIP1

(xi)SEQUENCE DESCRIPTION: SEQ ID NO: 33:

TTACTGGCTT	GGACCTCATT	GGCCATGACT	TGAGCTAAGA	TGCTAAGAGC	CCCAGCCAGG	60
TCATCTTGCT	CAGGTTTCATT	ATGGAGTCTA	GGGCAGACTC	TCACCTCCCT	GGACCATTTT	120

TAGAAATCTAT	GTGCCAGCTT	GCCAAAGACC	AACGAAAAAT	GCTTCTGGTG	GGGTCCAGGA	180
AGGCTGCGGA	GCAGGTGATA	CAAGACGCCC	TGAACCAGCT	TGAAGAACCT	CCTCTCATCA	240
GCTGCGCTGG	GTCTGCAGGT	ACACTTGCAA	TTGCCCAGCT	GGCAGGGGCC	AGGTCCCTTAC	300
AGCCTGAGAC	TCTGTTGATG	TTGAATCTCA	TGTGAGACTT	AGCTCAGGGG	CTCTCAGCCC	360
AGCAGCATGT	CAGCATTACC	TTAGGGGGCG	CCAGGCCCCA	TCCTAGATCA	GTTACATGTG	420
GAAACTCTGT	GCATTAGTGC	CTATACACTA	GTATTTTAGT	ATTTTCTT		468

## (2) INFORMATION FOR SEQ ID NO:34:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 393

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(x) FEATURE: exon 19 of HIP1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

CACTAGTAAG	CTCCTCCATT	CAGTGCTTAA	TTAACGAGGA	TGAAGCCAGC	TATGAGAACT	60
TGCTCTGACC	TTGCCCTGTG	TTCCCTCTCA	CAGATCACCT	CCTCTCCACG	GTCACATCCA	120
TTTCCAGCTG	CATCGAGCAA	CTGGAGAAAA	GCTGGAGCCA	GTATCTGGCC	TGCCCAGAAG	180
GTAAGAAATGG	CCAAGGACAG	TCTCTGTCGG	CTAGTGATGG	CCAGACAGGG	TTCAGAAGCA	240
CCTGAATGCG	GGGATAGTGA	CAGGTPCCCTC	TGCATCAAGA	AAGGCATGTA	GGCAACTCAT	300
ACAAGAAAGG	CATGTAGGCA	ACTCATAAAA	CGGGAGGAGA	GGGTATGAAA	GTGTCACCAT	360
CAACCAGACC	TGAGAAACTT	CTCTTTCCAA	TCC			393

## (2) INFORMATION FOR SEQ ID NO:35:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 421

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(x) FEATURE: exon 20 of HIP1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

GGCCTGCCCA	GAAGGTAAGA	ATGGCCAAGG	ACAGTCTCTG	TGGGCTAGTG	ATGGCCAGAC	60
AGGGTTTCAGA	AGCACCTGAA	TGCGGGGATA	GTGACAGGTC	CCTCTGCATC	AAGAAAGGCA	120
TGTAGGCAAC	TCATACAAGA	AAGGCATGTA	GGCAACTCAT	AAAACGGGAG	GAGAGGGTAT	180
GAAAGTGTCA	CCATCAACCA	GACCTGAGAA	ACTTCTCTTT	CCAATCCTGG	CAGACATCAG	240
TGGACTTCTC	CATTCCATAA	CCCTGCTGGC	CCACTTGACC	AGCGACGCCA	TTGCTCATGG	300
TGCCACCACC	TGCCCTCAGAG	CCCCACCTGA	GCCTGCCGAC	TGTGAGTACT	GGGGCATGAG	360
GGGCTGTTCA	TGGACCAGGG	GAGCAGGGGG	CCTTTAAAAG	TCTCTGTTGG	GCCGGGCGCA	420
G						421

## (2) INFORMATION FOR SEQ ID NO:36:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 498  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear  
 (ii) MOLECULE TYPE: genomic DNA  
 (iii) HYPOTHETICAL: no  
 (iv) ANTI-SENSE: no  
 (vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(x) FEATURE: exon 21 of HIP1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

AGGCCGAGGC	AGGAGAATCG	CTTGAACCTCA	GGAGGCGGAG	TTTGCAGTGA	GCCGAGATGG	60
CGCCACTGCA	CTCCAGCCTG	GGCAACAAGA	GCGAGACTCC	ATCTCAAAAA	AAAAGTGTCT	120
ATTGCCTTGT	ATCTCCAGCA	CTGACCGAGG	CCTGTAAGCA	GTATGGCAGG	GAAACCCTCG	180
CCTACCTGGC	CTCCCTGGAG	GAAGAGGGAA	GCCTTGAGAA	TGCCGACAGC	ACAGCCATGA	240
GGAAGTGCCT	GAGCAAGATC	AAGGCCATCG	GCGAGGTACT	TGGAGTAGTA	TCATTGAGGA	300
GCATTGTTAT	TCTTCTGGGT	GTGCGTGCTG	GTGAATGGCC	AGGGAATCGG	TGATGTTCTG	360
AGCTAGTTCT	TTCTGCACTT	AGAACTTGAT	TCTAGAAAGA	GATTGTTAAA	ATTGGAAAAT	420
CTGGCCGGGT	GCAGTGATTT	ATGCGTGTA	TCCCAGCACT	TTGGGAGGCC	GAGTCAGGAG	480
GATCACTTGA	GGCTAGAC					498

## (2) INFORMATION FOR SEQ ID NO:37:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 427  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear  
 (ii) MOLECULE TYPE: genomic DNA  
 (iii) HYPOTHETICAL: no  
 (iv) ANTI-SENSE: no  
 (vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(x) FEATURE: exon 22 of HIP1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

CCCTGTGGCT	TGCAGAAGGT	GTTTGCTGGG	TGGCCTCCTG	CCTTGCCATC	TTGTAAGGGT	60
TACAGATGGC	AGAGGAGAAG	AGACAGGAGG	CCCCAAGGTC	AGTTCAGCCT	TTGTGATGTG	120
TTCACAGGAG	CTCCTGCCCCA	GGGGACTGGA	CATCAAGCAG	GAGGAGCTGG	GGGACCTGGT	180
GGACAAGGAG	ATGGCGGCCA	CTTCAGCTGC	TATTGAAACT	GCCACGGCCA	GAATAGAGGT	240
AGGAGGTTCC	TGCAGGATCT	CCTGAAACGA	TGCCTTTGCA	GCTGCCCTTC	TGCAACACTG	300
CTCATTTAAAC	ATGTCACAGT	CGTTCATTAA	GGCCATGGCA	ACCCCTAAG	ACAGAAACCA	360
GAATTTGCCA	GGCACAGTGG	CTCATGCCTG	TAACCCAGC	ACCTTGGGAG	GATCACTTGA	420
GTCCAGG						427

## (2) INFORMATION FOR SEQ ID NO:38:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 367



- (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear  
 (ii) MOLECULE TYPE: genomic DNA  
 (iii) HYPOTHETICAL: no  
 (iv) ANTI-SENSE: no  
 (vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(x) FEATURE: exon 23 of HIP1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

CCCCCTGAAT	AGGTTAGAGT	CTGGATTCTT	TTCTGACTCT	CTCAAGAATG	TGGGCAGGGA	60
CTTGGGGACT	TCCAGATTCA	GGTTTCCCAG	CTACCACACG	ATGTTGGACT	GAAAGTATAG	120
TAAGACATTA	GTGGATCCTT	AATATTCAAG	GCACATTTAG	AAACCATGCT	TCTTTTTCAC	180
AGGAGATGCT	CAGCAAATCC	CGAGCAGGAG	ACACAGGAGT	CAAATTGGAG	GTGAATGAAA	240
GGTCGGTCTG	AGCGGCATGG	TGGGACCTAG	GGGAGCAGGA	TCTGTCTTCC	TGACATTGGT	300
CTATACTTTG	CATACTTATT	AGGGAATTAG	AGGAGAGCAG	TAGCAGCCAC	GGGGAAGGGC	360
TGAGTTG						367

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 502

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(x) FEATURE: exon 24 of HIP1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

CCCCGCAGAA	TGTTCCAGCA	ACCTCAGCAC	CCTTCTTACC	TCCCTTTCCC	ATTCCAAGCT	60
TGCCTTTGGC	TAGGAGTGGG	GAAGAGAACC	GTCGTGTTCA	TTGATCTTGG	ATCTTGATCT	120
CAGTGTATCC	TCGACTTGTT	TGTTTGGCAG	GATCCTTGGT	TGCTGTACCA	GCCTCATGCA	180
AGCTATTCAG	GTGCTCATCG	TGGCCTCTAA	GGACCTCCAG	AGAGAGATTG	TGGAGAGCGG	240
CAGGGTGAGC	GTGGGTGTGG	GCCCTGGGCA	GGAAGAGGAG	GCATCGGTGA	CAGACTCCCG	300
CTCCAACGGA	CTCTGTGATG	CTGCCGTCTT	ACTCTGTGTG	TCCACCTGAG	TACAGAGCAG	360
CCACTCCTGT	AGATATCAGC	AGAGGCCCTG	GGGAGAAGTC	AGAGCTCCAG	GACCTCCCCA	420
GAGGGTGGCC	AGGCATGTGT	CCCAACTCCA	GCTCCCTTCG	CACAGGCAGA	CATTGTGTGA	480
ACTTGCTGTG	GGAGCCCTTT	TT				502

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 437

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(x) FEATURE: exon 25 of HIP1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

TTTTGGTCTC	TGAATCTTCT	TCTTTTTTGT	AAAATGGGAA	TACTAATGCT	TATGTCTCAG	60
AGTTACTATG	AGGATGATTT	GGGATAATAT	ATGTATAAAA	GCACCTGCCA	TATAGTACAT	120
GCTCAATAAA	AGGTGGCTAT	TACTATTTT	TATTTCCCTA	GGGTACAGCA	TCCCCATAAG	180
AGTTTTATGC	CAAGAACTCT	CGATGGACAG	AAGGACTTAT	CTCAGCCTCC	AAGGCTGTGG	240
GCTGGGGAGC	CACGTGTCATG	GTGTAAGTAT	CTATTGGTAC	CAAGGGTCCT	CCCATGACCC	300
CTCTTCCATT	GATCCACTCC	AAACAATAGC	TAAGGAGGGA	AAAAAAAATC	TGTCCCTTAG	360
AAATAAATA	TTGATCAGGA	AGTCAATAGG	ACCGAGTTTA	CAAGGGAGCC	TGGCTCTCCC	420
AGGGGACACA	GGGCAGG					437

## (2) INFORMATION FOR SEQ ID NO:41:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 351

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(x) FEATURE: exon 26 of HIP1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

GGGAGCCTGG	CTCTCCCAGG	GGACACAGGG	CAGGCAGCCT	CCCCCCTG	TTTAGCCAAG	60
GGCGATGGGG	TGGTCTGGAG	GTGGGATTGT	GGAGGAGTTG	CAGCTCATTT	GCCCGTAACC	120
TAGTCCCTCT	TGTCGTTTTT	CATCAGGGAT	GCAGCTGATC	TGGTGGTACA	AGGCAGAGGG	180
AAATTTGAGG	AGCTAATGGT	GTGTTCTCAT	GAAATTGCTG	CTAGCACAGC	CCAGCTTGTG	240
GCTGCATCCA	AGGTAGGACC	TGGCTGGACC	TCCTAGGACG	CTGGAAGGCC	TGGTTAGAGA	300
GTACTAGGCT	AGGTTAAAGA	GTACTTGGCT	GCGTTAGGCA	GTACTTGGCT	G	351

## (2) INFORMATION FOR SEQ ID NO:42:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 418

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(x) FEATURE: exon 27 of HIP1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

CTTTTATAT	GATAGATATG	TCAGGAGCTG	ACTATAGTCA	GCAGATTTTG	AGAAGCTGAT	60
TGGTGATTGC	CGTTTGGCCC	ACATATGTTT	GCTAAGAACC	ATCAGAGCAA	TTATCTGATT	120
CAGTCCTTGT	TGCTCTAGGT	GTTGTATGAA	CCTAAATCTG	CTTTGTCCCTG	GTAGGTGAAA	180

GCTGATAAGG	ACAGCCCCAA	CCTAGCCCAG	CTGCAGCAGG	CCTCTCGGGG	AGTGAACCAG	240
GCCACTGCCG	GCGTTGTGGC	CTCAACCATT	TCCGGCAAAT	CACAGATCGA	AGAGACAGGT	300
AGCCTTTCCA	AAGGGACCCT	TTTCTTACCC	ACCCTGTTGA	GCTCTTCTCT	GCATCCTTCC	360
CTGTGATCCC	AACCAAATCC	CACAGGACTG	TGTCTAAATT	CTTTCATATT	TTTCATCT	418

## (2) INFORMATION FOR SEQ ID NO:43:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 279

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(x) FEATURE: exon 28 of HIP1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

TTTCCACAGA	GCATTGGCAT	TGGCTGCCTC	TCAGGTGCCA	GTCAGCCAGG	GTAGAATTTG	60
ATGAGACCTT	CTTGTTTCCA	TCCTTGCAGA	CAACATGGAC	TTCTCAAGCA	TGACGCTGAC	120
ACAGATCAAA	CGCCAAGAGA	TGGATTCTCA	GGTTAGGGTG	CTAGAGCTAG	AAAATGAATT	180
GCAGAAGGAG	CGTCAAAAAC	TGGGAGAGCT	TCGGAAAAAG	CACTACGAGC	TTGCTGGTGT	240
TGCTGAGGGC	TGGGAAGAAG	GTAAGCTGAC	TCAAAGGAT			279

## (2) INFORMATION FOR SEQ ID NO:44:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3715

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: genomic DNA

(iii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(x) FEATURE: exon 29 and partial cds of HIP1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

AACATAAATT	ATCATTGTCT	TTTAGGAACA	GAGGCATCTC	CACCTACACT	GCAAGAAGTG	60
GTAACCGAAA	AAGAATAGAG	CCAAACCAAC	ACCCCATATG	TCAGTGTAAG	TCCTTGTTAC	120
CTATCTCGTG	TGTGTTATTT	CCCCAGCCAC	AGGCCAAATC	CTTGGAGTCC	CAGGGGCAGC	180
CACACCACTG	CCATTACCCA	GTGCCGAGGA	CATGCATGAC	ACTTCCCAAA	GACTCCCTCC	240
ATAGCGACAC	CCTTTCTGTT	TGGACCCATG	GTCATCTCTG	TTCTTTTCCC	GCCTCCCTAG	300
TTAGCATCCA	GGCTGGCCAG	TGCTGCCCAT	GAGCAAGCCT	AGGTACGAAG	AGGGGTGGTG	360
GGGGGCAGGG	CCACTCAACA	GAGAGGACCA	ACATCCAGTC	CTGCTGACTA	TTTGACCCCC	420
ACAACAATGG	GTATCCTTAA	TAGAGGAGCT	GCTGTGTGTT	TGTTGACAGC	TTGGAAAGGG	480
AAGATCTTAT	GCCTTTTCTT	TTCTGTTTTC	TTCTCAGTCT	TTTCAGTTTC	ATCATTTGCA	540
CAAACTTG TG	AGCATCAGAG	GGCTGATGGA	TTCCAAACCA	GGACACTACC	CTGAGATCTG	600
CACAGTCAGA	AGGACGGCAG	GAGTGTCCCTG	GCTGTGAATG	CCAAAGCCAT	TCTCCCCCTC	660
TTTGGGCAGT	CCCATGGATT	TCCACTGCTT	CTTATGGTGG	TTGGTTGGGT	TTTTTGGTTT	720
TGTTTTTTTT	TTTAAAGTTT	CACTCACATA	GCCAACTCTC	CCAAAGGGCA	CACCCCTGGG	780
GCTGAGTCTC	CAGGGCCCCC	CAACTGTGGT	AGCTCCAGCG	ATGGTGCTGC	CCAGGCCTCT	840

CGGTGCTCCA	TCTCCGCCTC	CACACTGACC	AAGTGCTGGC	CCACCCAGTC	CATGCTCCAG	900
GCTCAGGCGG	AGCTGCTGAG	TGACAGCTTT	CCTCAAAAAG	CAGAAGGAGA	GTGAGTGCTT	960
TTCCCTCCTA	AAGCTGAATC	CCGGCGGAAA	GCCTCTGTCC	GCCTTTACAA	GGGAGAAGAC	1020
AACAGAAAGA	GGGACAAGAG	GGTTCACACA	GCCAGTTCC	CGTGACGAGG	CTCAAAAAC	1080
TGATCACATG	CTTGAATGGA	GCTGGTGAGA	TCAACAACAC	TACTTCCCTG	CCGGAATGAA	1140
CTGTCCGTGA	ATGGTCTCTG	TCAAGCGGGC	CGTCTCCCTT	GGCCCAGAGA	CGGATGTGG	1200
GAGTGATTCC	CAACTCCTTT	CTGCAGACGT	CTGCCTTGGC	ATCCTCTTGA	ATAGGAAGAT	1260
CGTTCCACTT	TCTACGCAAT	TGACAAACCC	GGAAGATCAG	ATGCAATTGC	TCCCATCAGG	1320
GAAGAACCCT	ATACTTGGTT	TGCTACCCCT	AGTATTTATT	ACTAACCTCC	CTTAAGCAGC	1380
AACAGCCTAC	AAAGAGATGC	TTGGAGCAAT	CAGAACTTCA	GGTGTGACTC	TAGCAAAGCT	1440
CATCTTTCTG	CCCGGCTACA	TCAGCCTTCA	AGAATCAGAA	GAAAGCCAAG	GTGCTGGACT	1500
GTTACTGACT	TGGATCCCAA	AGCAAGGAGA	TCATTTGGAG	CTCTTGGGTC	AGAGAAAATG	1560
AGAAAGGACA	GAGCCAGCGG	CTCCAACCTC	TTTCAGCCAC	ATGCCCCAGG	CTCTCGCTGC	1620
CCTGTGGACA	GGATGAGGAC	AGAGGGCACA	TGAACAGCTT	GCCAGGGATG	GGCAGCCCAA	1680
CAGCACTTTT	CCTCTTCTAG	ATGGACCCCA	GCATTTAAGT	GACCTTCTGA	TCTTGGGAAA	1740
ACAGCGTCTT	CCTTCTTTAT	CTATAGCAAC	TCATTTGGTG	TAGCCATCAA	GCACTTCCCA	1800
GGATCTGCTC	CAACAGAATA	TTGCTAGGTT	TTGCTACATG	ACGGGTGTGT	AGACTTCTGT	1860
TTGATCACTG	TGAACCAACC	CCCATCTCCC	TAGCCACCCC	CCCTCCCCAA	CTCCCTCTCT	1920
GTGCATTTTC	TAAGTGGGAC	ATTCAAAAAA	CTCTCTCCCA	GGACCTCGGA	TGACCATACT	1980
CAGACGTGTG	ACCTCCATAC	TGGGTTAAGG	AAGTATCAGC	ACTAGAAATT	GGGCAGTCTT	2040
AATGTTGAAT	GCTGCTTTCT	GCTTAGTATT	TTTTTGATTC	AAGGCTCAGA	AGGAATGGTG	2100
CGTGGCTTCC	CTGTCCCAGT	TGTGGCAACT	AAACCAATCG	GTGTGTTCTT	GATCGGGGTC	2160
AACATTTCCA	AAAGTGGCTA	GTCTCACTT	CTAGATCTCA	GCCATTCTAA	CTCATATGTT	2220
CCCAATTACC	AAGGGGTGGC	CGGGCACAGT	GGCTCACGCC	TGTAATCCCA	GCACTTTGAG	2280
AGGCTGAGGT	GGTAGGATCA	CCTGAGGTCA	GGAGTTCAAG	ACCAGCCTGT	CCAACATGGT	2340
GAAACCCCCA	TCTCTACTAA	AAATACCAAA	AATTAGCCGA	GCGTAGTGAC	GGGTGCCCCG	2400
AATCCAGCT	ACTCAGGAGG	CTGAGACAGG	AGAATCACCT	GAACCCAGAG	GGCAGAGGTT	2460
GCAGTGAGCT	GAGATCACGC	CATTGTACTC	CAGCCTGGGC	AACAAGAGCA	AAACTCCGTC	2520
TCAAAAAAAA	AAAAAATTA	CAAATGGGGC	AAACAGTCTA	GTGTAATGGA	TCAAATTAAG	2580
ATTCTCTGCC	CAGCCGGGCA	CAGTGGCGCA	TGCCTGTAAT	CCCAGAACTT	TGGGAGGCCA	2640
AGACGGGATG	ATTGCTTGAG	CTCAGGAGTT	TGAGACCAGG	CTGGGCATCA	TAGCAAGACC	2700
TCATCTCTAC	TAAAATTCAA	AAACAAAATT	AGCCGGGCAT	GATGGTGCAT	GCCTGTAGTC	2760
TCAGCTAGTT	GGGGAGCTAA	GGTGGGAGAA	TTGCTTGAGC	TTGGGAAGTC	GAGGCTGCAG	2820
TCAGCCCTGA	TTGTGCCAGT	GCACTCCGGC	CTGGGTGACA	GAGTGAGACC	CGTGCTCAAA	2880
AAAAAAAAGA	TTCTGTGTCA	GAGCCCAGCC	CAGGAGTTTG	AGGCTGCAAT	GAGCCATGAT	2940
TTCCCACTGC	ACTCCAGCCT	GAGTGACAGA	GCGAGACTCC	ATCTCTTTAA	AAACAAACAA	3000
AAAATTATCT	GAATGATCCT	GTCTCTAAAA	AGAAGCCACA	GAAATGTTTA	AAAACCTCAT	3060
CGACTTAGCC	TGAGTCATAA	CGGTAAAGAA	AGCACTTAAA	CAGAAGCAGA	GGCTAATTCA	3120
GTGTCACATG	AGGAAGTAGC	TGTCAGATGT	CACATAATTA	CTTTCGTAAT	AGCTCAGATT	3180
AGAATGGCTA	CCCCATTCTC	TAGACAAAAT	CAAATTGTCC	TATTGTGACT	CTTCTAAAAA	3240
TGAAGATGAA	GAGCTATTTA	ATGACACACC	TTGGATTAAA	ACGGGAATCA	CATCTTAAAG	3300
CTAAAAATGA	ACCTGCAAGC	CTTCTAAATG	AGTCACTGAG	CATCACTAGT	GACAAGTCTC	3360
GGGTGAGCGT	AAATGGGTCA	TGACAAGATG	GGACAGCAAC	AAAATCATGG	CTTAGGATCG	3420
ACAAGAAGTT	AAAAAACAGC	TGCATCTGTT	ACTTAAGTTT	GTAAGACAGT	GCCCTGAGAC	3480
CTCTAGAGAA	AAGATGTTTG	TTTACATAAG	AGAAAGAAGG	CCAGACATGG	TGTCTCACAC	3540
GTTTAATCCC	AGCACTTTGG	GAGGCAGGGG	CGGGTGGATC	ACCTGAGGTC	AGGAGTTCAA	3600
GACTAGCCTG	GCCAACATGG	TGAAACCCCG	TCTCTACTAA	AAATACAAA	ATTAGCCGGG	3660
CATGGTGGCA	GGCGCCTATA	ATCCCAGCTA	CTGGGGAGGC	TGAGGCAGGA	GAATC	3715